

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-313497

(43)Date of publication of application : 24.11.1998

(51)Int.Cl. H04R 3/00
G01S 5/18
G10L 3/02
G10L 7/04
G10L 9/00
H04S 7/00

(21)Application number : 09-252312

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>

(22)Date of filing : 17.09.1997

(72)Inventor : AOKI MARIKO
AOKI SHIGEAKI
MATSUI HIROYUKI
NISHINO YUTAKA

(30)Priority

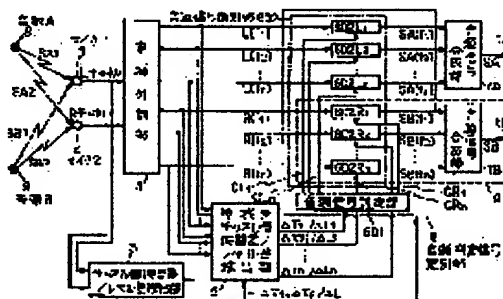
Priority number : 08246726 Priority date : 18.09.1996 Priority country : JP
09 76672 13.03.1997 JP
09 76695 13.03.1997 JP

(54) SOUND SOURCE SEPARATION METHOD, SYSTEM AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To separate sound accurately into respective components and to attain real time processing.

SOLUTION: A difference $\Delta \tau$ in time required for a sound signal to travel from a sound source to microphones 1, 2 is detected from output channel signals L, R of the microphones 1, 2, where the signals L, R are divided into frequency bands $L(f_1)-L(f_n)$, $R(f_1)-R(f_n)$ by Fourier transform, and an arrival time difference $\Delta \tau_i$ ($i=1, 2, \dots, n$) to the microphones 1, 2 of a corresponding frequency band of the $L(f_1)-L(f_n)$, $R(f_1)-R(f_n)$ and a signal level difference ΔL_i are detected. Then $L(f_1)-L(f_n)$, $R(f_1)-R(f_n)$ are divided into low frequencies $f_1 < 1/(2\Delta \tau)$, medium frequencies $1/(2\Delta \tau) < f_1 < 1/\Delta \tau$, and high frequencies $f_1 > 1/\Delta \tau$, and it is determined from which sound source the $L(f_i)$, $R(f_i)$ arrive, based on $\Delta \tau_i$, in the case of low frequencies, on ΔL_i , $\Delta \tau_i$ in the case of medium frequencies, and on ΔL_i in the case of high frequencies, and the result is outputted for each sound source and the output is subject to inverse Fourier transform by each sound source and the results are synthesized.



LEGAL STATUS

[Date of request for examination] 25.12.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(51) Int.Cl.⁸
 H 0 4 R 3/00
 G 0 1 S 5/18
 G 1 0 L 3/02
 7/04
 9/00

識別記号

3 2 0

3 0 1

F I

H 0 4 R 3/00

G 0 1 S 5/18

G 1 0 L 3/02

7/04

9/00

3 2 0

3 0 1 E

A

H

審査請求 未請求 請求項の数52 O L (全 33 頁) 最終頁に続く

(21) 出願番号 特願平9-252312

(22) 出願日 平成9年(1997)9月17日

(31) 優先権主張番号 特願平8-246726

(32) 優先日 平8(1996)9月18日

(33) 優先権主張国 日本 (J P)

(31) 優先権主張番号 特願平9-76672

(32) 優先日 平9(1997)3月13日

(33) 優先権主張国 日本 (J P)

(31) 優先権主張番号 特願平9-76695

(32) 優先日 平9(1997)3月13日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 青木 真理子

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(72) 発明者 青木 茂明

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(72) 発明者 松井 弘行

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(74) 代理人 弁理士 草野 卓

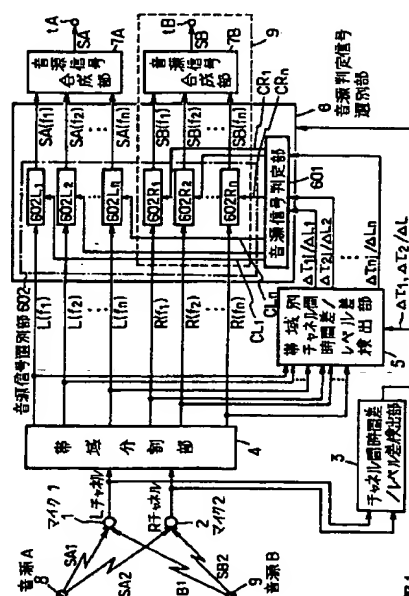
最終頁に続く

(54) 【発明の名称】 音源分離方法、装置及び記録媒体

(57) 【要約】

【課題】 正確に分離でき、かつ実時間処理を可能とする。

【解決手段】 マイクロホン1, 2の出力チャンネル信号 L, R から、音源からの音響信号がマイクロホン1, 2 に到達する時間差 $\Delta\tau$ を検出し、信号 L, R を、フーリエ変換により各周波数帯域 $L(f_1) \sim L(f_n)$, $R(f_1) \sim R(f_n)$ に分割し、 $L(f_1) \sim L(f_n)$ と $R(f_1) \sim R(f_n)$ の対応帯域のマイクロホン1, 2への到達時間差 $\Delta\tau_i$ ($i=1, 2, \dots, n$) と、信号レベル差 ΔL_i とを検出し、 $L(f_1) \sim L(f_n)$, $R(f_1) \sim R(f_n)$ を $f_i < 1/(2\Delta\tau)$ の低域と、 $1/(2\Delta\tau) < f_i < 1/\Delta\tau$ の中域と、 $f_i > 1/\Delta\tau$ の高域にわけ、低域では $\Delta\tau_i$ により、中域では ΔL_i と $\Delta\tau_i$ とにより、高域では ΔL_i により $L(f_i)$, $R(f_i)$ が何れの音源から到来したか判定して音源別に出力し、この出力を各音源別に逆フーリエ変換して合成する。



THIS PAGE BLANK (USPTO)

【特許請求の範囲】

【請求項1】 互いに離して設けられた複数のマイクロホンを用いて、複数の音源から少なくとも1つの音源を分離する音源分離方法であって、

上記各マイクロホンの各出力チャンネル信号を、複数の周波数帯域に分割する帯域分割過程と、

上記帯域分割過程で分割された各出力チャンネル信号の各同一帯域ごとに、上記複数のマイクロホンの位置に起因して変化する、マイクロホンに到達する音響信号のパラメータの値の差を、帯域別チャンネル間パラメータ値差として検出する帯域別チャンネル間パラメータ値差検出過程と、

上記各帯域の帯域別チャンネル間パラメータ値差にもとづき、その帯域の上記帯域分割された各出力チャンネル信号の何れがいずれの音源から入力された信号であるかを判定する音源信号判定過程と、

上記音源信号判定過程の判定にもとづき、上記帯域分割された各出力チャンネル信号から、同一音源から入力された信号を少なくとも1つ選択する音源信号選択過程と、上記音源信号選択過程で同一音源からの信号として選択された複数の帯域信号を音源信号として合成する音源合成過程とを有することを特徴とする音源分離方法。

【請求項2】 請求項1記載の方法において、上記帯域分割過程は各出力チャンネル信号の各分割された帯域信号は、主として1つの音源の音響信号の成分よりなる程度に、小さく分割することを特徴とする音源分離方法。

【請求項3】 請求項1又は2記載の方法において、上記帯域別チャンネル間パラメータ値差検出過程における上記パラメータ値は音源からの音響信号が上記各マイクロホンに到達するまでの時間であり、上記帯域別チャンネル間パラメータ値差は各マイクロホンに到達するまでの時間のマイクロホン間の差である帯域別チャンネル間時間差であることを特徴とする音源分離方法。

【請求項4】 請求項3記載の方法において、上記音響信号が各マイクロホンに到達するまでの時間のマイクロホン間の差をチャンネル間時間差として各マイクロホンの出力チャンネル信号から検出するチャンネル時間差検出過程を有し、

上記音源信号判定過程は、上記各帯域別チャンネル間時間差について、上記各チャンネル間時間差を照合して、その帯域の上記分割された各出力チャンネル信号がいずれの音源から入力された信号であるかを判定することを特徴とする音源分離方法。

【請求項5】 請求項4記載の方法において、上記チャンネル時間差検出過程は各出力チャンネル信号間の相互相関を求め、相互相関の各ピークとなるその出力チャンネル信号間の各時間差として上記各チャンネル間時間差を求めることを特徴とする音源分離方法。

【請求項6】 請求項5記載の方法において、

上記帯域別チャンネル間時間差は、上記各チャンネル間時間差中の、上記分割された各出力チャンネルの同一帯域の成分の位相差と対応する時間と最も近いものを求めて、その帯域別チャンネル間時間差とすることを特徴とする音源分離方法。

【請求項7】 請求項1又は2記載の方法において、上記帯域別チャンネル間パラメータ値差検出過程における上記パラメータ値は音源からの音響信号が上記各マイクロホンに到達した時の信号レベルであり、上記帯域別チャンネル間パラメータ値差は各分割された出力チャンネル信号の対応帯域間のレベル差である帯域別チャンネル間レベル差であることを特徴とする音源分離方法。

【請求項8】 請求項7記載の方法において、上記各マイクロホンの出力チャンネル信号間のレベル差を、チャンネル間レベル差として検出するチャンネル間レベル差検出過程と、

上記チャンネル間レベル差と、対応する帯域別チャンネル間レベル差の全てと比較する比較過程と、

その比較過程で分割帯域の所定数以上が同様の関係にあれば、上記チャンネル間レベル差にもとづき、対応する出力チャンネル信号の全帯域について同一の音源から入力された信号であると判定し、上記比較過程で所定値以上が同様の関係になれば、上記帯域別にいずれの音源から入力された信号であるかを判定する上記音源信号判定過程を実行することを特徴とする音源分離方法。

【請求項9】 請求項1又は2記載の方法において、上記パラメータ値は音源からの音響信号が上記マイクロホンに到達するまでの時間と、その音響信号が到達した時の信号レベルであり、上記帯域別チャンネル間パラメータ値差として帯域別チャンネル間時間差と、帯域別チャンネル間レベル差が求められ、

各音源からの音響信号が上記各マイクロホンに到達するまでの時間のマイクロホン間の差を、各マイクロホンの出力チャンネル信号から、チャンネル時間差として検出するチャンネル間時間差検出過程と、

上記チャンネル間時間差を基準にして上記分割された各出力チャンネル信号を、低域、中域、高域の3つの周波数領域に分け領域分割過程とを有し、

上記音源信号判定過程は、

上記分割された低域の周波数帯域については、上記帯域別チャンネル間時間差を利用して対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する過程と、

上記分割された中域の周波数帯域については、上記帯域別チャンネル間レベル差と、上記帯域別チャンネル間時間差を利用して、対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する過程と、

上記分割された高域の周波数帯域については、上記帯域別チャンネル間レベル差を利用して、対応する帯域の分割

された各出力チャンネル信号の何れかがいずれの音源からの入力信号であるか判定する過程とからなることを特徴とする音源分離方法。

【請求項10】 請求項1～9の何れかに記載の方法において、

上記帯域別チャンネル間パラメータ値差検出過程において、その互いに差をとるべき、もとのチャンネル信号の周波数帯域が異なる場合は、その周波数帯域が互いに重ならない周波数帯域は、上記帯域別チャンネル間パラメータ値差検出過程を実行せず、上記音源信号判定過程ではその信号がある帯域を予め知られている広い帯域の音源からの入力信号と判定することを特徴とする音源分離方法。

【請求項11】 互いに離して設けられた複数のマイクロホンを用いて、複数の音源から少なくとも1つの音源を分離する音源分離方法であって、

上記各マイクロホンの各出力チャンネル信号のパワースペクトルを求めるスペクトル分解過程と、

上記各チャンネルごとのパワースペクトルを、主としてほぼ1つの音源の成分が含まれるように複数の周波数帯域に分割する帯域分割過程と、

上記各同一帯域ごとに、各チャンネル間で分割されたパワースペクトル差を、帯域別チャンネル間レベル差として検出する帯域別チャンネル間レベル差検出過程と、

上記各帯域の帯域別チャンネル間レベル差にもとづき、その帯域の信号が上記出力チャンネル信号の何れであるかを判定する音源信号判定過程と、

上記音源信号判定過程の判定にもとづき、上記分割されたパワースペクトルから、同一音源からの信号を少なくとも1つ選択する音源信号選択過程と、

上記音源信号選択過程で同一音源からのものとして選択されたスペクトルを音源信号として合成する音源合成過程とを有することを特徴とする音源分離方法。

【請求項12】 請求項11記載の方法において、

上記各マイクロホンの出力チャンネル信号間のレベル差をチャンネル間レベル差として検出するチャンネル間レベル差検出過程と、

上記チャンネル間レベル差と、対応する上記帯域別チャンネル間レベル差の全てとを比較する比較過程と、

その比較過程で分割帯域の所定数以上が同様の関係であれば、上記チャンネル間レベル差にもとづき、対応する出力チャンネル信号の全帯域について同一音源から入力された信号であると判定し、上記比較過程で所定値以上が同様の関係になれば、上記音源信号判定過程を実行することを特徴とする音源分離方法。

【請求項13】 請求項1乃至12の何れかの方法において、

上記各マイクロホンの出力チャンネル信号を、各帯域が主として1つの音源信号成分になる程度に、複数の周波数帯域に分割する第2帯域分割過程と、

上記第2帯域分割過程で分割された各出力チャンネル信号の帯域別レベルをそれぞれ検出する帯域別レベル検出過程と、

その帯域別レベル検出過程で検出された各帯域別レベルを同一帯域についてチャンネル間で比較した結果にもとづき発音をしていない音源を検出する音源状態判定過程と、

その音源状態判定過程で得た発音をしていない音源の検出信号により、上記音源合成過程で合成された音源信号のうち、上記発音をしていない音源と対応する合成信号を抑圧する信号抑圧過程とを有することを特徴とする音源分離方法。

【請求項14】 請求項13の方法において、

上記音源状態判定過程は、上記各帯域別レベルのチャンネル間での比較で、最も大きいチャンネルを帯域ごとに決定する過程と、

各チャンネルごとに最もレベルが大きい帯域の数を求める過程と、

上記最もレベルが大きい帯域の数が第1基準値を越えるか否かを判定する第1判定過程と、

その第1判定過程で第1基準値を越えると判定すると、その越えた最もレベルが大きい帯域の数と対応するチャンネルのマイクロホン位置から、発音している1個の音源を推定する過程と、

その推定された音源以外の音源を発音していないものとして検出する過程とを有することを特徴とする音源分離方法。

【請求項15】 請求項1～12の何れかに記載の方法において、

上記各マイクロホンの各出力チャンネル信号のパワースペクトルを求めるスペクトル分解過程と、

上記各チャンネルごとのパワースペクトルを、主としてほぼ1つの音源の成分が含まれるように周波数帯域を分割して帯域別レベルをそれぞれ検出する帯域別レベル検出過程と、

これら各帯域別レベルを同一帯域について比較し、最大レベルのチャンネルを各帯域ごとに決定する過程と、

各チャンネルごとの最大レベルの帯域の数を求める過程と、

その帯域の数が第1基準値を越えたか否かを判定する第1判定過程と、

その第1判定過程で第1基準値を越える数と判定すると、その越えたチャンネルのマイクロホンが受けもつ、ゾーンから発音している1個の音源を推定する過程と、

その推定された音源以外の音源は発音していないと判定する過程と、

上記音源合成過程で合成された音源信号のうち、上記発音していないと判定された音源と対応する信号を、抑圧する信号抑圧過程とを有することを特徴とする音源分離方法。

【請求項16】 請求項14又は15の方法において、上記第1判定過程で、第1基準値を超えるものがないと判定されると、上記最もレベルが大きい帯域の数が、上記第1基準値よりも小さい第2基準値以下か否かを判定する第2判定過程と、

その第2判定過程で、第2基準値より小さいと判定されると、その小さいと判定された最もレベルが大きい帯域の数と対応するチャンネルのマイクロホン位置から、発言していない1個の音源として検出する過程とを有することを特徴とする音源分離方法。

【請求項17】 請求項1乃至12の何れかの方法において、

上記各マイクロホンの出力チャンネル信号を、各帯域が主として1つの音源信号成分になる程度に、複数の周波数帯域に分割する第2帯域分割過程と、

上記第2帯域分離過程で分割された各出力チャンネル信号のそのマイクロホンへの到達時間差を同一帯域ごとに検出する帯域別時間差検出過程と、

この帯域別時間差検出過程で検出された各帯域別到達時間差を、同一帯域についてチャンネル間で比較した結果にもとづき発言をしていない音源を検出する音源状態判定過程と、

その音源状態判定過程で得た発言をしていない音源の検出信号により、上記音源合成過程で合成された音源信号のうち、上記発言をしていない音源と対応する合成信号を抑圧する信号抑圧過程とを有することを特徴とする音源分離方法。

【請求項18】 請求項3の方法において、上記帯域別チャンネル間時間差を、同一帯域についてチャンネル間で比較した結果にもとづき発言をしていない音源を検出する音源状態判定過程と、

その音源状態判定過程で得た発言をしていない音源を検出信号により、上記音源合成過程で合成された音源信号のうち、上記発言をしていない音源と対応する信号を抑圧する信号抑圧過程とを有することを特徴とする音源分離方法。

【請求項19】 請求項17又は18の方法において、上記音源状態判定過程は、上記各帯域別到達時間差比較で最も速く音源信号が到達したチャンネルを帯域ごとに決定する過程と、

各チャンネルごとに最も速く到達した帯域の数が第1基準値を超えるか否かを判定する第1判定過程と、

その第1判定過程が第1基準値を超えると判定すると、その越えた最も速く到達した帯域数と対応するチャンネルのマイクロホン位置から発言している1個の音源を推定する過程と、

その推定された音源以外の音源を発言していないものとして検出する過程とを有することを特徴とする音源分離方法。

【請求項20】 請求項18の方法において、

上記第1判定過程で、第2基準値を超えるものがないと判定されると、上記最も速く到達する帯域の数が、上記第1基準値よりも小さい第2基準値より小さいか否かを判定する第2判定過程と、

その第2判定過程で、第2基準値より小さいと判定されると、その小さいと判定された最も速い到達時間の帯域数と対応するチャンネルのマイクロホン位置から、発言していない1個の音源として検出する過程とを有することを特徴とする音源分離方法。

10 【請求項21】 請求項16又は20の方法において、音源が4個以上の場合で、上記第3判定過程で、第3基準値より小さいと判定されると、上記第2基準値を上記第1基準値を超えない範囲内で、順次大きくして、上記第2判定過程と同じ判定を、(M-2)回以内、Mは音源の数、繰返す過程を有することを特徴とする音源分離方法。

【請求項22】 請求項13～21の何れかに記載の方法において、

各出力チャンネル信号の全周波数成分のレベルをそれぞれ検出する全帯域レベル検出過程と、

20 その全帯域レベル検出過程で検出した各チャンネルの全周波数成分レベルの何れもが第3基準値以下であるかを判定し、何れかが第3基準値以下でないと判定すると上記音源状態判定過程に移る第3判定過程とを有することを特徴とする音源分離方法。

【請求項23】 請求項22の方法において、

上記第3判定過程が第3基準値以下であると判定されると、上記音源合成過程で合成された各音源信号のすべてを抑圧する過程を有することを特徴とする音源分離方法。

【請求項24】 請求項13～23の何れかの方法において、

上記帯域分割過程と上記第2帯域分割過程は同一過程として行われることを特徴とする音源分離方法。

【請求項25】 互いに離して設けられた複数のマイクロホンを用いて、複数の音源から少なくとも1つの音源を分離する音源分離装置であって、

上記各マイクロホンの各出力チャンネル信号を、主として1つの音源の音響信号の成分のみが含まれる程度に複数の周波数帯域に分割すると共に、これら分割された各出力チャンネル信号の各同一帯域ごとに、上記複数のマイクロホンの位置に起因して変化する、マイクロホンに到達する音響信号のパラメータの値の差を、帯域別チャンネル間パラメータ値差として検出する帯域別チャンネル間パラメータ値差検出手段と、

上記各帯域の帯域別チャンネル間パラメータ値差にもとづき、その帯域の上記帯域分割された各出力チャンネル信号の何れがいずれの音源から入力された信号であるかを判定する音源信号判定手段と、

50 上記音源信号判定過程の判定にもとづき、上記帯域分割

された各出力チャンネル信号から、同一音源から入力された信号を少なくとも1つ選択する音源信号選択手段と、上記音源信号選択過程で同一音源からの信号として選択された、複数の帯域信号を音源信号として合成する音源合成手段とを具備することを特徴とする音源分離装置。

【請求項26】 請求項25の装置において、上記帯域別チャンネル間パラメータ値差検出手段における上記パラメータ値は音源からの音響信号が上記各マイクロホンに到達するまでの時間であり、上記帯域別チャンネル間パラメータ値差は各マイクロホンに到達するまでの時間のマイクロホン間の差である帯域別チャンネル間時間差であることを特徴とする音源分離装置。

【請求項27】 請求項25の装置において、上記音響信号が各マイクロホンに到達するまでの時間のマイクロホン間の差をチャンネル間時間差として各マイクロホンの出力チャンネル信号から検出するチャンネル時間差検出手段を有し、

上記音源信号判定手段は、上記各帯域別チャンネル間時間差について、上記各チャンネル間時間差を照合して、その帯域の上記分割された各出力チャンネル信号がいずれの音源から入力された信号であるかを判定する手段であることを特徴とする音源分離装置。

【請求項28】 請求項25の装置において、上記帯域別チャンネル間パラメータ値差検出手段における上記パラメータ値は音源からの音響信号が上記各マイクロホンに到達した時の信号レベルであり、上記帯域別チャンネル間パラメータ値差は各分割された出力チャンネル信号の対応帯域間のレベル差である帯域別チャンネル間レベル差であることを特徴とする音源分離装置。

【請求項29】 請求項28の装置において、上記各マイクロホンの出力チャンネル信号間のレベル差を、チャンネル間レベル差として検出するチャンネル間レベル差検出手段と、

上記チャンネル間レベル差と、対応する帯域別チャンネル間レベル差の全てと比較する比較手段と、その比較手段で分割帯域の所定数以上が同様の関係にあれば、上記チャンネル間レベル差にもとづき、対応する出力チャンネル信号の全帯域について同一の音源から入力された信号であると判定し、上記比較手段で所定値以上が同様の関係になれば、上記帯域別にいずれの音源から入力された信号であるかを判定する上記音源信号判定手段を実行する手段を含むことを特徴とする音源分離装置。

【請求項30】 請求項25の装置において、上記パラメータ値は音源からの音響信号が上記マイクロホンに到達するまでの時間と、その音響信号が到達した時の信号レベルであり、上記帯域別チャンネル間パラメータ値差として帯域別チャンネル間時間差と、帯域別チャンネル間レベル差が求められ、

各音源からの音響信号が上記各マイクロホンに到達するまでの時間のマイクロホン間の差と、各マイクロホンの

出力チャンネル信号から、チャンネル時間差として検出するチャンネル間時間差検出手段と、

上記チャンネル間時間差を基準にして、上記分割された各出力チャンネル信号を、低域、中域、高域の3つの周波数領域に分ける領域分割手段とを有し、

上記音源信号判定手段は、

上記分割された低域の周波数帯域については、上記帯域別チャンネル間時間差を利用して対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する手段と、

上記分割された中域の周波数帯域については、上記帯域別チャンネル間レベル差と、上記帯域別チャンネル間時間差を利用して、対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する手段と、

上記分割された高域の周波数帯域については、上記帯域別チャンネル間レベル差を利用して、対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する手段とからなることを特徴とする音源分離装置。

【請求項31】 請求項25乃至30の何れかの装置において、

上記帯域分割された各出力チャンネル信号の帯域別レベルをそれぞれ検出する帯域別レベル検出手段と、

その帯域別レベル検出手段が検出された各帯域別レベルを同一帯域についてチャンネル間で比較した結果にもとづき発音をしていない音源を検出する音源状態判定手段と、

その音源状態判定手段で得た発音をしていない音源の検出信号により、上記音源合成手段で合成された音源信号のうち、上記発音をしていない音源と対応する信号を抑圧する信号抑圧手段とを有することを特徴とする音源分離装置。

【請求項32】 請求項31の装置において、

上記音源状態判定手段は、上記各帯域別レベルのチャンネル間での比較で、最も大きいチャンネルを帯域ごとに決定する手段と、

各チャンネルごとに最もレベルが大きい帯域の数を求める手段と、

上記最もレベルが大きい帯域の数が第1基準値を越えるか否かを判定する第1判定手段と、

その第1判定手段で第1基準値を越えると判定すると、その越えた最もレベルが大きい帯域の数と対応するチャンネルのマイクロホン位置から、発音している1個の音源を推定する手段と、

その推定された音源以外の音源を発音していないものとして検出する手段とを有することを特徴とする音源分離装置。

【請求項33】 請求項31の装置において、

上記第1判定手段で、第1基準値を越えるものがないと

判定されると、上記最もレベルが大きい帯域の数が、上記第1基準値よりも小さい第2基準値以下か否かを判定する第2判定手段と、

その第2判定手段で、第2基準値より小さいと判定されると、その小さいと判定された最もレベルが大きい帯域の数と対応するチャンネルのマイクロホン位置から、発音していない1個の音源として検出する手段とを有することを特徴とする音源分離装置。

【請求項34】 請求項25乃至30の何れかの装置において、

上記帯域分割された各出力チャンネル信号のそのマイクロホンへの到達時間差を同一帯域ごと検出する帯域別時間差検出手段と、

この帯域別時間差検出手段で検出された各帯域別到達時間差を、同一帯域についてチャンネル間で比較した結果にもとづき発音をしていない音源を検出する音源状態判定手段と、

その音源状態判定手段で得た発音をしていない音源を検出信号により、上記音源合成手段で合成された音源信号のうち、上記発音をしていない音源と対応する信号を抑圧する信号抑圧手段とを有することを特徴とする音源分離装置。

【請求項35】 請求項34の装置において、上記音源状態判定手段は、上記各帯域別到達時間差比較で最も速く音源信号が到達したチャンネルを帯域ごとに決定する手段と、

各チャンネルごとに最も速く到達した帯域の数が第1基準値を越えるか否かを判定する第1判定手段と、

その第1判定手段が第2基準値を越えると判定すると、その越えた最も速く到達した帯域数と対応するチャンネルのマイクロホン位置から発音している1個の音源を推定する手段と、

その推定された音源以外の音源を発音をしていないものとして検出する手段とを有することを特徴とする音源分離装置。

【請求項36】 請求項35の装置において、上記第1判定手段で、第1基準値を越えるものがないと判定されると、上記最も速く到達する帯域の数が、上記第1基準値よりも小さい第2基準値以下か否かを判定する第2判定手段と、

その第2判定手段で、第2基準値より小さいと判定されると、その小さいと判定された最も速い到達時間の帯域数と対応するチャンネルのマイクロホン位置から、発音していない1個の音源として検出する手段とを有することを特徴とする音源分離装置。

【請求項37】 請求項31～36の何れかに記載の装置において、

各出力チャンネル信号の全周波数成分のレベルをそれぞれ検出する全帯域レベル検出手段と、

その全帯域レベル検出手段で検出した各チャンネルの全周

波数成分レベルの何れもが第3基準値以下であるかを判定し、何れかが第1基準値以下でないと判定すると、上記音源状態判定手段に移る第3判定手段とを有することを特徴とする音源分離装置。

【請求項38】 互いに離して設けられた複数のマイクロホンを用いて、複数の音源から少なくとも1つの音源を分離する下記過程を有する音源分離方法のプログラムを記録した記録媒体であって、

上記各マイクロホンの各出力チャンネル信号を、主に1つの音源の音響信号の成分のみを含む程度に複数の周波数帯域に分割すると共にこれら分割された各出力チャンネル信号の各同一帯域ごとに、上記複数のマイクロホンの位置に起因して変化する、マイクロホンに到達する音響信号のパラメータの値の差を、帯域別チャンネル間パラメータ値差として検出する帯域別チャンネル間パラメータ値差検出過程と、

上記各帯域の帯域別チャンネル間パラメータ値差にもとづき、その帯域の上記帯域分割された各出力チャンネル信号の何れがいずれの音源から入力された信号であるかを判定する音源信号判定過程と、

上記音源信号判定過程の判定にもとづき、上記帯域分割された各出力チャンネル信号から、同一音源から入力された信号を少なくとも1つ選択する音源信号選択過程と、上記音源信号選択過程で同一音源からの信号として選択された複数の帯域信号を音源信号として合成する音源合成過程とを有するコンピュータにより読出し可能な記録媒体。

【請求項39】 請求項38の記録媒体において、上記帯域別チャンネル間パラメータ値差検出過程における上記パラメータ値は音源からの音響信号が上記各マイクロホンに到達するまでの時間であり、上記帯域別チャンネル間パラメータ値差は各マイクロホンに到達するまでの時間のマイクロホン間の差である帯域別チャンネル間時間差であって、

上記プログラムは上記音響信号が各マイクロホンに到達するまでの時間のマイクロホン間の差をチャンネル間時間差として各マイクロホンの出力チャンネル信号から検出するチャンネル時間差検出過程を有し、

上記音源信号判定過程は、上記各帯域別チャンネル間時間差について、上記各チャンネル間時間差を照合して、その帯域の上記分割された各出力チャンネル信号がいずれの音源から入力された信号であるかを判定することを特徴とする記録媒体。

【請求項40】 請求項39の記録媒体において、上記チャンネル時間差検出過程は各出力チャンネル信号間の相互相関を求め、相互相関の各ピークとなる、その出力チャンネル信号間の各時間差として上記各チャンネル間時間差を求めることを特徴とする記録媒体。

【請求項41】 請求項40の記録媒体において、上記帯域別チャンネル間時間差は、上記各チャンネル間時間

差中の、上記分割された各出力チャンネルの同一帯域の成分の位相差と対応する時間と最も近いものを求めて、その帯域別チャンネル間時間差とすることを特徴とする記録媒体。

【請求項42】 請求項38の記録媒体において、上記帯域別チャンネル間パラメータ値差検出過程における上記パラメータ値は音源からの音響信号が上記各マイクロホンに到達した時の信号レベルであり、上記帯域別チャンネル間パラメータ値差は各分割された出力チャンネル信号の対応帯域間のレベル差である帯域別チャンネル間レベル差であって、
上記プログラムは上記各マイクロホンの出力チャンネル信号間のレベル差を、チャンネル間レベル差として検出するチャンネル間レベル差検出過程と、
上記チャンネル間レベル差と、対応する帯域別チャンネル間レベル差の全てと比較する比較過程と、その比較過程で分割帯域の所定数以上が同様の関係にあれば、上記チャンネル間レベル差にもとづき、対応する出力チャンネル信号の全帯域について同一の音源から入力された信号であると判定し、上記比較過程で所定値以上が同様の関係になれば、上記帯域別にいずれの音源から入力された信号であるかを判定する、上記音源信号判定過程と実行する過程とを有することを特徴とする記録媒体。

【請求項43】 請求項38の記録媒体において、上記プログラムは上記パラメータ値は音源からの音響信号が上記マイクロホンに到達するまでの時間と、その音響信号が到達した時の信号レベルであり、上記帯域別チャンネル間パラメータ値差として帯域別チャンネル間時間差と、帯域別チャンネル間レベル差が求められ、各音源からの音響信号が上記各マイクロホンに到達するまでの時間のマイクロホン間の差を、各マイクロホンの出力チャンネル信号から、チャンネル時間差として検出するチャンネル間時間差検出過程と、
上記チャンネル間時間差を基準にして、上記分割された各出力チャンネル信号を、低域、中域、高域の3つの周波数領域に分ける領域分割過程とを有し、
上記音源信号判定過程は、
上記分割された低域の周波数帯域については、上記帯域別チャンネル間時間差を利用して対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する過程と、
上記分割された中域の周波数帯域については、上記帯域別チャンネル間レベル差と、上記帯域別チャンネル間時間差を利用して、対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する過程と、
上記分割された高域の周波数帯域については、上記帯域別チャンネル間レベル差を利用して、対応する帯域の分割された各出力チャンネル信号の何れがいずれの音源からの入力信号であるかを判定する過程とからなることを特徴と

する記録媒体。

【請求項44】 請求項38乃至43の何れかの記録媒体において、

上記プログラムは上記帯域分割された各出力チャンネル信号の帯域別レベルをそれぞれ検出する帯域別レベル検出過程と、

その帯域別レベル検出過程が検出された各帯域別レベルを同一帯域についてチャンネル間で比較した結果にもとづき発音をしていない音源を検出する音源状態判定過程と、

その音源状態判定過程で得た発音をしていない音源の検出信号により、上記音源合成過程で合成された音源信号のうち、上記発音をしていない音源と対応する信号を抑圧する信号抑圧過程とを有することを特徴とする記録媒体。

【請求項45】 請求項44の記録媒体において、上記音源状態判定過程は、上記各帯域別レベルのチャンネル間での比較で、最も大きいチャンネルを帯域ごとに決定する過程と、

各チャンネルごとに最もレベルが大きい帯域の数を求める過程と、

上記最もレベルが大きい帯域の数が第1基準値を越えるか否かを判定する第1判定過程と、

その第1判定過程で第1基準値を越えると判定すると、その越えた最もレベルが大きい帯域の数と対応するチャンネルのマイクロホン位置から、発音している1個の音源を推定する過程と、

その推定された音源以外の音源を発音をしていないものとして検出する過程とを有することを特徴とする記録媒体。

【請求項46】 請求項45の記録媒体において、上記プログラムは上記第1判定過程で、第1基準値を越えるものがないと判定されると、上記最もレベルが大きい帯域の数が、上記第1基準値よりも小さい第2基準値以下か否かを判定する第3判定過程と、

その第2判定過程で、第2基準値より小さいと判定されると、その小さいと判定された最もレベルが大きい帯域の数と対応するチャンネルのマイクロホン位置から、発音をしていない1個の音源として検出する過程とを有することを特徴とする記録媒体。

【請求項47】 請求項38乃至43の何れかの記録媒体において、

上記プログラムは上記帯域分割された各出力チャンネル信号のそのマイクロホンへの到達時間差を同一帯域ごと検出する帯域別時間差検出過程と、

この帯域別時間差検出過程で検出された各帯域別到達時間差を、同一帯域についてチャンネル間で比較した結果にもとづき、発音をしていない音源を検出する音源状態判定過程と、

その音源状態判定過程で得た発音をしていない音源を検

出した検出信号により、上記音源合成過程で合成された音源信号のうち、上記発音していない音源と対応する信号を抑圧する信号抑圧過程とを有することを特徴とする記録媒体。

【請求項48】 請求項47の記録媒体において、上記音源状態判定過程は、上記各帯域別到達時間差比較で、最も速く音源信号が到達したチャンネルを帯域ごとに決定する過程と、各チャンネルごとに最も速く到達した帯域の数が第1基準値を越えるか否かを判定する第1判定過程と、その第1判定過程が第2基準値を越えると判定すると、その越えた最も速く到達した帯域数と対応するチャンネルのマイクロホン位置から発音している1個の音源を推定する過程と、その推定された音源以外の音源を発音していないものとして検出する過程とを有することを特徴とする記録媒体。

【請求項49】 請求項48の記録媒体において、上記プログラムは上記第1判定過程で、第1基準値を越えるものがないと判定されると、上記最も速く到達する帯域の数が、上記第1基準値よりも小さい第2基準値より小さいか否かを判定する第2判定過程と、その第3判定過程で、第2基準値より小さいと判定されると、その小さいと判定された最も速い到達時間の帯域数と対応するチャンネルのマイクロホン位置から、発音していない1個の音源として検出する過程とを有することを特徴とする記録媒体。

【請求項50】 請求項46又は49の記録媒体において、音源が4個以上の場合で、上記プログラムは上記第2判定過程で、第2基準値より小さいと判定されると、上記第2基準値を上記第1基準値を越えない範囲内で、順次大きくして、上記第2判定過程と同じ判定を(M-2)以内、Mは音源の数、繰返す過程を有することを特徴とする記録媒体。

【請求項51】 請求項44～50の何れかに記載の記録媒体において、上記プログラムは各出力チャンネル信号の全周波数成分のレベルをそれぞれ検出する全帯域レベル検出過程と、その全帯域レベル検出過程で検出した各チャンネルの全周波数成分レベルの何れもが第3基準値以下であるかを判定し、何れかが第3基準値以下でないと判定すると、上記音源状態判定過程に移る第3判定過程とを有することを特徴とする記録媒体。

【請求項52】 請求項47～50の何れかに記載の記録媒体において、上記プログラムは上記チャンネル時間差検出過程は上記帯域時間差検出過程を兼ねていることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は音声信号源や各種環境音源などの複数の音源から発せられた複数の音響信号が混ざった信号から少なくとも1つの音源の信号を分離抽出する方法、その方法に用いた音源分離装置、およびその方法をコンピュータにより実行するためのプログラムを記録した記録媒体に関する。

【0002】この種の音源分離装置は、例えばテレビ会議における收音装置、騒音環境下で発声した音声信号の伝送のための收音装置、音源の種類を識別する装置の收音装置など各種のものに適用される。従来の音源分離技術は、周波数領域において各信号の基本周波数を推定し、調波構造を抜き出すことにより、同一音源からの成分を集めて合成する方法が用いられてきた。

【0003】しかしこの方法では、(1)分離可能な信号が、音声の母音や楽音のような調波構造を持つものに限定されるという問題があった、(2)基本周波数の推定は一般に長い処理時間を必要とするため、実時間で音源を分離することは困難であった、(3)調波構造の推定誤りなどにより、抽出された信号に他の音源の周波数成分が混じり、それが雑音として知覚されるため分離精度が不十分であった。

【0004】

【発明の解決しようとする課題】この発明の目的は調波構造を持たない音源の音響信号でも分離抽出することとし、つまり音源の種類に依存することなく音源分離を可能とし、かつ実時間で音源分離を可能とする方法、装置、及びプログラム記録媒体を提供することにある。

【0005】この発明の他の目的は分離精度が高く、雑音の混入が少ない音源分離方法、装置及びプログラム記録媒体を提供することにある。

【0006】

【課題を解決するための手段】この発明の音源分離方法は互いに離して設けられた複数のマイクロホンを用い、上記各マイクロホンの各出力チャンネル信号を、帯域分割過程で複数の周波数帯域に分割し、その各帯域には主として1つの音源信号成分のみ存在するようにし、これら分割された各出力チャンネル信号の各同一帯域ごとに、上記複数のマイクロホンの位置に起因して変化する、マイクロホンに到達する音響信号のパラメータ、つまりレベル(パワー)、到達時間の値の差を、帯域別チャンネル間パラメータ値差として検出し、上記各帯域の帯域別チャンネル間パラメータ値差にもとづき、その帯域の上記帯域分割された各出力チャンネル信号の何れがいずれの音源から入力された信号であるかを音源信号判定過程で判定し、この音源信号判定過程の判定にもとづき、上記帯域分割された各出力チャンネル信号から、同一音源から入力された信号を少なくとも1つ、音源信号選択過程で選択し、その音源信号選択過程で同一音源からの信号として選択された、複数の帯域信号を音源信号として音源合成過程で合成する。

【0007】この発明の音源分離方法の実施例によれば、上記帯域分割過程で分割された各出力チャンネル信号の帯域別レベルをそれぞれ検出し、これらが検出された各帯域別レベルを同一帯域についてチャンネル間で比較した結果にもとづき発音をしていない音源を検出し、その発音をしていない音源の検出信号により、上記音源合成過程で合成された音源信号のうち、上記発音をしていない音源と対応する合成信号を抑圧する。

【0008】この発明の音源分離方法の他の実施例によると、上記帯域分離過程で分割された各出力チャンネル信号のそのマイクロホンへの到達時間差を同一帯域ごとに検出し、これら検出された各帯域別到達時間差を、同一帯域についてチャンネル間で比較した結果にもとづき発音をしていない音源を検出し、その発音をしていない音源の検出信号により、上記音源合成過程で合成された音源信号のうち、上記発音をしていない音源と対応する合成信号を抑圧する。

【0009】

【発明の実施の形態】図1にこの発明の実施例を示す。マイクロホン1、2が間隔、例えば20cm程度をあけて配され、これらマイクロホン1、2はそれぞれ音源A、Bからの音響信号を収集して電気信号に変換する。マイクロホン1の出力をLチャンネル信号と、マイクロホン2の出力をRチャンネル信号と称する。Lチャンネル信号とRチャンネル信号はチャンネル間時間差／レベル差検出部3と、帯域分割部4へ供給され、帯域分割部4ではそれぞれ複数の周波数帯域信号に分割されて帯域別チャンネル間時間差／レベル差検出部5と音源判定信号選別部6へ供給される。検出部3、5の各検出出力に応じて選別部6において各帯域ごとに何れかのチャンネル信号がA成分又はB成分として選別され、これら選別された帯域ごとのA成分信号、B成分信号はそれぞれ音源信号合成部7A、7Bでそれぞれ合成されて、音源A信号と音源B信号とに分離出力される。音源Aがマイクロホン2よりマ*

$$\Delta \tau_1 = 1000 \times \Delta \alpha_1 / F \quad (1)$$

$$\Delta \tau_2 = 1000 \times \Delta \alpha_2 / F \quad (2)$$

ただしFはサンプリング周波数であり、1000倍にするのは演算の便宜上値をある程度大きくするためである。時間差 $\Delta \tau_1$ 、 $\Delta \tau_2$ は、音源A、Bそれぞれの信号のLチャンネル信号とRチャンネル信号のチャンネル間時間差である。

【0012】図1、2の説明に戻って帯域分割部4はLチャンネル信号とRチャンネル信号をそれぞれ各周波数帯域の信号 $L(f_1)$ 、 $L(f_2)$ 、 \dots 、 (f_n) と、信号 $R(f_1)$ 、 $R(f_2)$ 、 \dots 、 (f_n) に分割する(S04)。この分割は例えば各チャンネル信号をそれぞれ離散的フーリエ変換して周波数領域信号に変換した後、各周波数帯域に分割することにより行う。この帯域分割は、音源A、Bの各信号の周波数特性の差から各帯域において、一方の音源の信号成分のみが主として存在する

*マイクロホン1に近いと、音源Aよりマイクロホン1に到達する信号SA1は音源Aよりマイクロホン2に到達する信号SA2より早く到達し、かつレベルが大きい、また音源Bがマイクロホン1よりマイクロホン2に近いと、音源Bからマイクロホン1、2にそれぞれ到達する信号SB1、SB2は後者が早くマイクロホン2に到達し、レベルも大きい。このようにこの発明では、音源のマイクロホン1、2に対する位置に起因する両マイクロホン1、2に到達する音響信号の変化量、この例では両信号の到達時間差とレベル差を利用する。

【0010】図1に示した装置は以下に示すように動作する。図2に示すように、マイクロホン1、2に2つの音源A、Bからの信号が取り込まれる(S01)。チャンネル間時間差／レベル差検出部3は、Lチャンネル信号とRチャンネル信号からチャンネル間時間差またはレベル差を検出する。時間差の検出に用いるパラメータとしては、Lチャンネル信号とRチャンネル信号との相互相関関数を用いた場合で説明する。図3に示すようにまず、Lチャンネル信号とRチャンネル信号との各サンプル $L(t)$ 、 $R(t)$ を読み込み(S02)、これらサンプル間の相互相関関数を算出する(S03)。この算出は両チャンネル信号が同一サンプル時点についての相互相関を求め、また一方のチャンネル信号に対し他方のチャンネル信号をサンプル時点を1つだけずらした場合、2つだけずらした場合・・・の各場合の相互相関をそれぞれ求めて相互相関関数を求める。これら相互相関を多数求め、これらをパワーで正規化したヒストグラムを作成する(S04)。次に、ヒストグラムの累積度数順位第一位、第二位をそれぞれとる時点差 $\Delta \alpha_1$ 、 $\Delta \alpha_2$ を求める(S05)。これらの時点差 $\Delta \alpha_1$ 、 $\Delta \alpha_2$ を、次式によりそれぞれチャンネル間時間差 $\Delta \tau_1$ 、 $\Delta \tau_2$ に変換して出力する(S06)。

【0011】

程度、音声信号の場合は、例えば20Hz帯域幅で分割する。音源Aのパワースペクトルが例えば図4Aに示すように得られ、音源Bのパワースペクトルが図4Bに示すように得られ、この各スペクトルが分離できる程度の帯域幅 Δf で分割する。この時、例えば破線で対応するスペクトルを示すように、一方の音源のスペクトルに対し他方の音源のスペクトルは無視できる。またこの図4A、4Bから理解されるように帯域幅 $2\Delta f$ で分離してもよい。つまり、各帯域に1本のスペクトルのみが含まれるようにしなくてもよい。なお、離散的フーリエ変換は例えば20～40msごとに行う。

【0013】次に、帯域別チャンネル間時間差／レベル差検出部5は、例えば $L(f_1)$ と $R(f_1)$ 、 \dots 、 $L(f_n)$ と $R(f_n)$ といった各対応する帯域信号のチャネ

ル間について、帯域別チャンネル間時間差またはレベル差を検出する(S05)。ここで、帯域別チャンネル間時間差は、チャンネル間時間差検出部3で検出したチャンネル間*

$$\Delta\tau_1 - \{(\Delta\phi_i / (2\pi f_i)) + (k_{i1} / f_i)\} = \varepsilon_{i1} \quad (3)$$

$$\Delta\tau_2 - \{(\Delta\phi_i / (2\pi f_i)) + (k_{i2} / f_i)\} = \varepsilon_{i2} \quad (4)$$

$i=1, 2, \dots, n$, $\Delta\phi_i$ は信号L(f_i)と信号R(f_i)の位相差である。これら式で ε_{i1} , ε_{i2} が最小になるように整数 k_{i1} , k_{i2} を決める。次に、その最小値の ε_{i1} と ε_{i2} とを比べて小さい方のチャンネル時間差 $\Delta\tau_j$ ($j=1, 2$)を、その帯域 i のチャンネル間時間差 $\Delta\tau_{ij}$ とする。つまり一方の音源信号のその帯域でのチャンネル間時間差とする。

【0015】音源判定信号選別部6は、帯域別チャンネル間時間差/レベル差検出部5で検出された帯域別チャンネル間時間差 $\Delta\tau_{ij} \sim \tau_{nj}$ を用いて各帯域信号L(f_1) \sim L(f_n)とR(f_1) \sim R(f_n)との各対応するものについて何れを選択するか判定を音源信号判定部601で行う(S06)。例えば、チャンネル間時間差/レベル差検出部3で算出された時間差 $\Delta\tau_1$, $\Delta\tau_2$ のうち、 $\Delta\tau_1$ が、L側のマイクロホンに近い、音源Aからの信号のチャンネル間時間差であり、 $\Delta\tau_2$ が、R側のマイクロホンに近い、音源Bからの信号のチャンネル間時間差である場合で説明する。

【0016】この場合、帯域別チャンネル間時間差/レベル差検出部5で算出された時間差 $\Delta\tau_{ij}$ が $\Delta\tau_1$ である帯域 i は、音源信号判定部601によりゲート602L $_i$ が開とされてL側の入力信号L(f_i)がそのままSA(f_i)として出力され、R側の帯域 i の入力信号R(f_i)は音源信号判定部601によりゲート602Rが閉とされてSB(f_i)は0として出力される。時間差 $\Delta\tau_{ij}$ が $\Delta\tau_2$ となる帯域 i は、逆に、L側は信号L(f_i)はSA(f_i)=0として出力され、R側は入力信号R(f_i)がそのままSB(f_i)として出力される。つまり図1に示すように帯域信号L(f_1) \sim L(f_n)はそれぞれゲート602L $_1 \sim 602L_n$ を通じて音源信号合成部7Aへ供給され、帯域信号R(f_1) \sim R(f_n)はそれぞれゲート602R $_1 \sim 602R_n$ を通じて音源信号合成部7へ供給される。音源判定信号選別部6内の音源信号判定部601では $\Delta\tau_{1j} \sim \Delta\tau_{nj}$ が入力され、 $\Delta\tau_{ij}$ が $\Delta\tau_1$ と判定された帯域 i についてはゲート制御信号CL $_i=1$ とCR $_i=0$ が生成され、対応するゲート602L $_i$ が開、602R $_i$ が閉にそれぞれ制御され、 $\Delta\tau_{ij}$ が $\Delta\tau_2$ と判定された帯域 i についてはゲート制御信号CL $_i=0$ と、CR $_i=1$ が生成され、対応するゲート602L $_i$ が閉、602R $_i$ が開にそれぞれ制御される。以上の説明は機能構成であって、実際には例えばデジタルシグナルプロセッサにより処理される。

【0017】音源信号合成部7Aで信号SA(f_i) \sim SA(f_n)が合成され、前記帯域分割の例ではそれぞ

*時間差 $\Delta\tau_1$, $\Delta\tau_2$ を利用することにより一意的に検出される。この検出に用いる式は以下のとおりである。

【0014】

れ逆フーリエ変換され、信号SAとして出力端子 t_A に出力され、また音源信号合成部7Bで信号SB(f_i) \sim SB(f_n)が同様に合成されて信号SBとして出力端子 t_B に出力される。以上の説明で明らかのように、この発明装置においては、各チャンネル信号の細かく帯域分割した、各帯域成分がそれぞれどの音源からのものであるかを判定し、判定された成分は全て出力する、すなわち、音源A, Bの信号の周波数成分が互いに重なっていなければ、特定の周波数帯域を欠落させることなく処理を行うため、調波構造のみ抜き出す従来の方法に比べて音質を高く保ったまま音源A, Bの各信号を分離することが可能である。

【0018】以上の説明は、チャンネル間時間差/レベル差検出部3及び帯域別チャンネル間時間差/レベル差検出部5で検出した、チャンネル間時間差と、帯域別チャンネル間時間差のみを利用して、音源判定信号部601で判定条件を決定した。次にこの判定条件の決定をチャンネル間のレベル差を用いて処理する実施例を説明する。この実施例は図5に示すようにマイクロホン1, 2からLチャンネル信号とRチャンネル信号を取込み(S02)、これらLチャンネル信号とRチャンネル信号のチャンネル間レベル差 ΔL をチャンネル間時間差/レベル差検出部3(図1)で検出する(S03)。図2中のステップS04と同様に、Lチャンネル信号、Rチャンネル信号をそれぞれ n 個の帯域別チャンネル信号L(f_1) \sim L(f_n), R(f_1) \sim R(f_n)に分割し(S04)、帯域別チャンネル信号L(f_1) \sim L(f_n)とR(f_1) \sim R(f_n)との対応帯域、つまりL(f_1)とR(f_1), L(f_2)とR(f_2), ..., L(f_n)とR(f_n)について帯域別チャンネル間レベル差 ΔL_1 , ΔL_2 , ..., ΔL_n を検出する(S05)。

【0019】人間の音声は、20ms \sim 40ms程度の間は定常状態とみなすことが出来る。そのため、音源信号判定部601(図1)においては、20ms \sim 40ms毎に、チャンネル間レベル差 ΔL の対数を取った値の符号と、帯域別チャンネル間レベル差 ΔL_i の対数を取った値の符号とが、全帯域のうち何割以上の帯域で、同じ符号(+又は-)になるのかを算出し、所定値、例えば8割以上の帯域で両者が同じ符号を持てば(S06, S07)、そこから20ms \sim 40msの間はチャンネル間レベル差 ΔL のみで判定し(S08)、同じ符号を持つのが8割以下の帯域であれば、そこから20ms \sim 40msの間は帯域毎に、帯域別チャンネル間レベル差 ΔL_i を用いて判定する(S09)。判定の仕方は、全帯域をチャンネル間レベル差 ΔL で判定する場合は、 ΔL が正であ

れば、Lチャンネル信号 $L(t)$ がそのまま信号SAとして出力され、Rチャンネル信号 $R(t)$ は信号SB=0として出力される。 ΔL が0以下であれば逆に、Lチャンネル信号 $L(t)$ は信号SA=0として出力され、Rチャンネル信号 $R(t)$ がそのまま信号SBとして出力される。ただし、これは、チャンネル間レベル差としてL側からR側を引いた値を用いた場合の説明である。また、帯域別チャンネル間レベル差 ΔL_i を用いて帯域毎に判定する場合は、各帯域 f_i ごとに帯域別チャンネル間レベル差 ΔL_i が正であれば、L側分割信号 $L(f_i)$ がそのまま信号SA(f_i)として出力され、R側分割信号R(f_i)は信号SB(f_i)=0として出力される。レベル差 ΔL_i が0以下であれば逆に、L側は分割信号L(f_i)は信号SA(f_i)=0として出力され、R側は分割信号R(f_i)が信号SB(f_i)として出力される。以上のようにして音源信号判定部601からゲート制御信号 $CL1 \sim CLn$, $CR1 \sim CRn$ が出力され、ゲート602L1 \sim 602Ln, 602R1 \sim 602Rnがそれぞれ制御される。これも、前者と同様、帯域別チャンネル間レベル差として、L側からR側を引いた値を用いた場合の説明である。信号SA(f_1) \sim SA(f_n)、信号SB(f_1) \sim SB(f_n)は先の実施例と同様にそれぞれ合成された信号SA, SBとして出力端子 t_A , t_B にそれぞれ出力される(S10)。

【0020】前記実施例では、音源信号判定部601で用いる判定条件として、到達時間差とレベル差のうちどちらかの片方のみを利用する。しかし、レベル差のみを利用した場合、低域の周波数帯域では $L(f_i)$ とR(f_i)とのレベルが拮抗する場合があります、その場合はレベル差を正確に求めることが困難になる。また、時間差のみを利用した場合は、高い周波数帯域においては、位相の回転が起こるため時間差を正しく算出することが困難な場合がある。これらの点から、低域の周波数帯域では時間差を、高域ではレベル差を判定に用いた方が、全帯域に渡り単一のパラメータを用いるよりも有利である場合がある。

【0021】そこで、音源信号判定部601で帯域別チャンネル間時間差と帯域別チャンネル間レベル差を共に用いる実施例を図6以下の図面を参照して説明する。この実施例の機能構成のブロックとしては図1と同一であるが、チャンネル間時間差/レベル差検出部分3、帯域別チャンネル間時間差/レベル差検出部5と音源信号判定部601での処理が以下のように異なる。チャンネル間時間差/レベル差検出部3は、検出された時間差 $\Delta \tau_1$, $\Delta \tau_2$ の各絶対値の平均、又は $\Delta \tau_1$, $\Delta \tau_2$ が比較的近い*

$$\Delta \sigma_i = 1000 \cdot \Delta \phi_i / 2\pi f_i$$

分割された信号 $L(f_i)$, $R(f_i)$ が中域と判定された場合は図8に示すように帯域別チャンネル間レベル差 $\Delta L(f_i)$ を利用して、位相差 $\Delta \phi_i$ を一意に決定する。即ち $\Delta L(f_i)$ が正かを調べ(S23)、正であ

*値であれば、その一方のみなど、一つの時間差 $\Delta \tau$ を出力する。なおチャンネル間時間差 $\Delta \tau_1$, $\Delta \tau_2$, $\Delta \tau$ をチャンネル信号 $L(t)$, $R(t)$ を周波数軸上で帯域分割する前に算出したが、帯域分割した後に算出することも可能である。

【0022】図5に示すように、Lチャンネル信号 $L(t)$ 、Rチャンネル信号 $R(t)$ をフレーム(例えば20 \sim 40ms)毎に読み込み(S02)、帯域分割部4でLチャンネル信号、Rチャンネル信号をそれぞれ複数の周波数帯域に分割する。この例ではLチャンネル信号 $L(t)$ 、Rチャンネル信号 $R(t)$ にそれぞれハニング窓をかけ(S03)、それぞれフーリエ変換を施して分割された信号 $L(f_1) \sim L(f_n)$ 、 $R(f_1) \sim R(f_n)$ を得る(S04)。

【0023】次に、帯域別チャンネル間時間差/レベル差検出部5では分割された信号の周波数 f_i が $1/(2\Delta \tau)$ ($\Delta \tau$ はチャンネル時間差)以下の帯域(以下、低域と呼ぶ)であるかを調べ(S05)、以下であれば帯域別チャンネル間位相差 $\Delta \phi_i$ を出力し(S08)、分割された信号の周波数 f が $1/(2\Delta \tau)$ より大きく $1/\Delta \tau$ 未満の帯域(以下、中域と呼ぶ)であるかがチェックされ(S06)、この中域であれば帯域別チャンネル間位相差 $\Delta \phi_i$ 及びレベル差 ΔL_i を出力し(S09)、分割された信号の周波数 f が $1/\Delta \tau$ 以上の帯域(以下、高域と呼ぶ)かがチェックされ(S07)、高域であれば帯域別チャンネル間レベル差 ΔL_i を出力する(S10)。

【0024】音源信号判定部601は、帯域別チャンネル間時間差/レベル差検出部5で検出された帯域別チャンネル間位相差、レベル差を用いて $L(f_1) \sim L(f_n)$ 、 $R(f_1) \sim R(f_n)$ それぞれについて何れを出力するかの判定を行う。なお、位相差 $\Delta \phi_i$ 、レベル差 ΔL_i については、この例では共にL側からR側の値を引いて算出した値を用いる。

【0025】低域と判定された信号 $L(f_i)$, $R(f_i)$ については図7に示すようにまず位相差 $\Delta \phi_i$ が π 以上かを調べ(S15)、 π 以上であれば $\Delta \phi_i$ から 2π を減算した値を $\Delta \phi_i$ とし(S17)、ステップS15で $\Delta \phi_i$ が π 以上でなければ、 $-\pi$ 以下かを調べ(S16)、以下であれば $\Delta \phi_i$ に 2π を加算した値を $\Delta \phi_i$ とし(S18)、ステップS16で $-\pi$ 以下でなければ $\Delta \phi_i$ をそのまま用いる(S19)。ステップS17, S18, S19で求めた帯域別チャンネル間位相差 $\Delta \phi_i$ を時間差 $\Delta \sigma_i$ に次式で変換する(S20)。

【0026】

$$(5)$$

れば、その帯域別チャンネル間位相差 $\Delta \phi_i$ が正であるかを調べ(S24)、正であればその $\Delta \phi_i$ をそのまま出力し(S26)、ステップS24で正でなければ $\Delta \phi_i$ に 2π を加算した値を $\Delta \phi_i$ として出力する(S2

7)。ステップS23で $\Delta L(f_i)$ が正でなければ、その帯域別チャンネル間位相差 $\Delta \phi_i$ が負であるかを調べ(S25)、負であれば、その $\Delta \phi_i$ をそのまま $\Delta \phi_i$ として出力し(S28)、ステップS25で負でなければ $\Delta \phi_i$ から 2π を減算した値を $\Delta \phi_i$ として出力する*

$$\Delta \phi_i = 1000 \cdot \Delta \phi_i / 2\pi f_i$$

以上のようにして低域、中域における帯域別チャンネル間時間差 $\Delta \sigma_i$ と、高域における帯域別チャンネル間レベル差 $\Delta L(f_i)$ が得られ、これらに応じて音源信号の判別が次のようになされる。図9に示すように低域と中域においては位相差 $\Delta \phi_i$ を、高域においてはレベル差 ΔL_i を利用して両チャンネルの各周波数成分を該当するどちらかの音源の信号として判別する。具体的には、低域と中域においては図7、8でそれぞれ求められた帯域別チャンネル間時間差 $\Delta \sigma_i$ が正であるかを調べ(S34)、正であれば、その帯域iのL側チャンネル信号 $L(f_i)$ を信号 $SA(f_i)$ として出力し、R側帯域チャンネル信号 $R(f_i)$ を0の信号 $SB(f_i)$ として出力する(S36)。ステップS34で帯域別チャンネル時間差 $\Delta \sigma_i$ が正でない場合は逆に $SA(f_i)$ として0を出力し、 $SB(f_i)$ としてR側チャンネル信号 $R(f_i)$ を出力する(S37)。

【0028】また、高域においては、図6中のステップS10で検出した帯域別チャンネル間レベル差 $\Delta L(f_i)$ が正であるかを調べ(S35)、正であれば信号 $SA(f_i)$ としてL側チャンネル信号 $L(f_i)$ を出力し、 $SB(f_i)$ として0を出力する(S38)。ステップS35でレベル差 ΔL_i が正でなければ $SA(f_i)$ として0を出力し、 $SB(f_i)$ としてR側帯域チャンネル信号 $R(f_i)$ を出力する(S39)。

【0029】以上のようにして各帯域についてL側又はR側が出力され、音源信号合成部7A、7Bでそれぞれ判別した各周波数成分を全帯域に渡り加算し(S40)、かつ、加算した各信号を逆フーリエ変換し(S41)、その変換した信号 SA 、 SB を出力する(S42)。以上説明したように、この実施例においては、周波数帯域毎に音源分離に有利なパラメータを用いることにより、全帯域に渡り単一のパラメータを用いる場合に比べてより分離性能の高い音源分離を実現することが可能である。

【0030】この発明は音源の数が3個以上でも適用できる。例として、音源数が3、マイクロホン数が2である場合でマイクロホンへの到達時間差を利用して音源分離する場合を説明する。この場合、チャンネル間時間差/レベル差検出部3で各音源についてLチャンネル信号、Rチャンネル信号のチャンネル間時間差を算出する際に、図3に示したように相互相関のパワーで正規化したヒストグラムの、累積度数(ピーク値)第一位から第三位までをとる各時点を求めることによって各音源信号についてのチャンネル間時間差 $\Delta \tau_1$ 、 $\Delta \tau_2$ 、 $\Delta \tau_3$ を算出する。

* (S29)。これらステップS26～S29の何れかの $\Delta \phi_i$ が次式によりその帯域別チャンネル間時間差 $\Delta \sigma_i$ として演算される(S30)。

【0027】

(6)

そして、帯域別チャンネル間時間差/レベル差検出部5においても、各帯域の帯域別チャンネル間時間差を $\Delta \tau_1$ から $\Delta \tau_3$ のどれかに決定する。この決定の仕方は、前記実施例で述べた計算式(3)、(4)と同様である。音源信号判定部601では、例として、 $\Delta \tau_1 > 0$ 、 $\Delta \tau_2 > 0$ 、 $\Delta \tau_3 < 0$ である場合で説明する。ここで、 $\Delta \tau_1$ 、 $\Delta \tau_2$ 、 $\Delta \tau_3$ はそれぞれ、音源A、B、C各信号のチャンネル間時間差と仮定し、さらに、これらの値はL側からR側の値を引いて算出した値と仮定する。この場合、音源AはL側のマイクロホン1に近く、音源BはR側のマイクロホン2の近くにある。よって、Lチャンネルの信号から、帯域別チャンネル間時間差が $\Delta \tau_1$ となる帯域の信号を加算して音源Aの信号を、また $\Delta \tau_2$ となる帯域を加算して、音源Bの信号をそれぞれ分離することが可能である。また、Rチャンネル信号から、帯域別チャンネル間時間差が $\Delta \tau_3$ となる帯域の信号を加算して出力することにより、音源Cの信号を分離する。

【0031】上述では音源信号を分離し、分離された各音源信号 SA 、 SB を各別に出力した。しかし、例えば一方の音源Aは発話者による音声であり、他方の音源Bは騒音のような場合、騒音と混合された音源Aの信号音を分離抽出し、騒音を抑圧するためにもこの発明を適用することができる。その場合は図1において音源信号合成部7Aを残し、1点鎖線で示す枠9中の音源信号合成部7B、ゲート602R1～602Rnを省略すればよい。

【0032】一方の音源Aが他方の音源Bより周波数帯域が広い場合でその各周波数帯域が予め知られている場合は、図10に示すように図1において帯域分離部10において、両音源信号の重なっていない周波数帯域を分離する。例えば音源Aの信号 $A(t)$ の周波数帯域は $f_1 \sim f_n$ であるが音源Bの信号 $B(t)$ の周波数帯域は $f_1 \sim f_n$ ($f_n > f_m$)の場合、重なっていない帯域 $f_{m+1} \sim f_n$ の信号をマイクロホン1、2の出力から分離し、この帯域 $f_{m+1} \sim f_n$ の信号については、音源信号判定部601の判定処理、場合によっては帯域別チャンネル間時間差/レベル差検出部5の処理を行わず、音源信号判定部601は、音源Bの信号として選出するチャンネル信号 $SB(t)$ として選出するRの分割された帯域チャンネル信号 $R(f_{m+1}) \sim R(f_n)$ をそれぞれ $SB(f_{m+1}) \sim SB(f_n)$ として出力し、 $SA(f_{m+1}) \sim SA(f_n)$ は0を出力させるように音源信号選択部602を制御する。即ちゲート602Lm+1～602Lnは常閉とし、ゲート602Rm+1～

602Rnは常開とする。

【0033】上述では各帯域別チャンネル間時間差 $\Delta\sigma_i$ 、正か負かにより、また各帯域別チャンネル間レベル差 ΔL_i が正か負かにより、つまり、いずれも0をしきい値として、その帯域信号が何れのマイクロホンに近いかを判別した。これはマイクロホン1として結ぶ線の2等分線に対して音源Aと音源Bと左右対称に位置している場合である。この関係にない場合は判別しきい値を以下のように決めればよい。

【0034】音源Aの信号がマイクロホン1、マイクロホン2に到達する帯域別チャンネル間レベル差を ΔL_A 、到達する帯域別チャンネル間時間差を $\Delta\tau_A$ 、音源Bの信号がマイクロホン1、マイクロホン2に到達する帯域別チャンネル間レベル差を ΔL_B 、到達する帯域別チャンネル間時間差を $\Delta\tau_B$ とそれぞれする。このとき、帯域別チャンネル間レベル差のしきい値 ΔL_{th} は

$$\Delta L_{th} = (\Delta L_A + \Delta L_i) / 2$$

とし、帯域別チャンネル間時間差のしきい値 $\Delta\tau_{th}$ は

$$\Delta\tau_{th} = (\Delta\tau_A + \Delta\tau_B) / 2$$

とすればよい。先に述べた実施例では $\Delta L_B = -\Delta L_A$ 、 $\Delta\tau_B = -\Delta\tau_A$ の場合で $\Delta L_{th} = 0$ 、 $\Delta\tau_{th} = 0$ となる。音源A、Bを分離できるように、二つの音源をマイクロホン1、2に対し、互いに異なる側となるように、マイクロホン1、2を位置させ、マイクロホン1、2に対する距離、方向は必ずしも正しくはわかっていない場合があり、しきい値 ΔL_{th} 、 $\Delta\tau_{th}$ を可変として、分離がよく行われるように ΔL_{th} 、 $\Delta\tau_{th}$ を調整可能としてもよい。

【0035】前記実施例では部屋の残響や回折の影響により、帯域別チャンネル間時間差や帯域別チャンネル間レベル差に誤りが生じ、各音源信号を精度よく分離することができない場合がある。このような問題を改善した実施例を次に述べる。図11に示すように、マイクロホンM1、M2、M3は、例えば1辺が20cmの正三角形の頂点の位置に配置されている。マイクロホンM1～M3の指向特性に基づいて空間が分割して設定され、その各分割された空間を音源ゾーンと呼ぶ。全てのマイクロホンM1～M3が無指向で同じ特性を有する場合には、例えば図12に示すように、ゾーンZ1～Z6のように6個に分割される。つまり、各マイクロホンM1、M2、M3と、その中心点Cpをそれぞれ通る直線により、中心点Cpを中心に等角間隔で6分割された6つのゾーンZ1～Z6が形成される。音源AはゾーンZ3に、音源BはゾーンZ4に位置している。つまり、1個の音源ゾーンには1個の音源が属するよう、マイクロホンM1～M3の配置や特性に基づいて各音源ゾーンを決定する。

【0036】図11において、帯域分割部41は、マイクロホンM1で收音した第1チャンネルの音響信号S1をn個の周波数帯域信号S1(f1)～S1(fn)に分割し、分割部42でマイクロホンM2で收音した第2チ

ャネルの音響信号S2をn個の周波数帯域信号S2(f1)～S2(fn)に分割し、帯域分割部43は、マイクロホンM3で收音した第3チャンネルの音響信号S3をn個の周波数帯域信号S3(f1)～S3(fn)に分割する。これら各帯域f1～fnは帯域分割部41～43で共通であり、このような帯域分割は離散的フーリエ変換器を利用することができる。

【0037】音源分離部80は図1乃至図10を参照して説明した手法を用いて音源信号を分離するものである。ただし図11ではマイクロホンが3つであるから、この3つのチャンネルの信号の各2つの組合せについて同様な処理を行う。従って音源分離部80内の帯域分割部と帯域分割部41～43を兼用することもできる。帯域別レベル(パワー)検出部S1で帯域分割部41で得られた各帯域の信号S1(f1)～S1(fn)のレベル(パワー)信号P(S1f1)～P(S1fn)が検出され、同様に帯域別レベル検出部52、53でそれぞれ帯域分割部42、43で得られた各帯域信号S2(f1)～S2(fn)、S3(f1)～S3(fn)の各P(S2f1)～P(S2fn)、P(S3f1)～P(S3fn)がそれぞれ検出される。これら帯域別レベル検出もフーリエ変換器で実現できる。つまり各チャンネル信号を離散的フーリエ変換によりスペクトルに分解し、その各スペクトルの電力を求めればよい。従って、各チャンネル信号について、パワースペクトルを求め、そのパワースペクトルを帯域分割してもよい。各マイクロホンM1～M3の各チャンネル信号を、帯域別レベル検出部400で各帯域に分割すると共にそのレベル(パワー)を出力することになる。

【0038】一方全帯域レベル検出部61でマイクロホンM1で收音された第1チャンネルの音響信号S1の全周波数成分のレベル(パワー)P(S1)が検出され、全帯域レベル検出部62、63でそれぞれマイクロホンM2、M3でそれぞれ收音された第2、第3チャンネル2、3の各音響信号S2、S3の全周波数成分のレベルP(S2)、P(S3)が検出される。

【0039】音源状態判定部70では、コンピュータ処理により、音響を発していない音源ゾーンを判定する。まず、帯域別レベル検出部50により得られる帯域別レベルP(S1f1)～P(S1fn)、P(S2f1)～P(S2fn)、P(S3f1)～P(S3fn)を、同一の帯域の信号について相互に比較する。そして各帯域f1～fn毎に、最も大きなレベルのチャンネルを特定する。

【0040】帯域分割の数nを所定数以上にすることにより、前述したように、1つの帯域には1個の音源の音響信号しか含まれないと見なせるようにすることができるので、同一帯域fiのレベルP(S1fi)、P(S2fi)、P(S3fi)は、同一音源からの音響のレベルと見なすことができる。よって、第1～第3チ

チャンネルについて同一の帯域のレベル $P(S1fi)$, $P(S2fi)$, $P(S3fi)$ に差があるときは、音源に最も近いマイクロホンのチャンネルの帯域のレベルが最も大きくなる。

【0041】前記処理の結果、各帯域 $f1 \sim fn$ について、最もレベルの大きなチャンネルがそれぞれ割り当てられる。 n 個の帯域中で第1～第3各チャンネルについて、最もレベルが大きな帯域の合計数 $\chi1$, $\chi2$, $\chi3$ を算出する。この合計数の値が大きいチャンネルのマイクロホンほど、音源に近いとみなすことができる。合計数値が例えば $90n/100$ 以上程度であればそのチャンネルのマイクロホンに音源が近いと判定することができる。しかし、最もレベルが大きい帯域の合計数が $53n/100$ 、次に合計値が大きい値が $49n/100$ の場合はそのそれぞれの対応マイクロホンに音源が近いと明確ではない。従って当該合計数が予め設定した基準値 ThP 、例えば $n/3$ 程度を越えたとき、当該合計数と対応するチャンネルのマイクロホンにその音源が最も近いと判定する。

【0042】また、この音源状態判定部70には、全帯域レベル検出部60で検出された各チャンネルのレベル $P(S1) \sim P(S3)$ も入力されていて、そのレベルの全てが予め設定した基準値 ThR 以下の場合には、何れのゾーンにも、音源がないと判定する。この音源状態判定部70による判定結果に基づき、制御信号を発生して、音源分離部80で分割された音響信号A, Bに対する抑圧を信号抑圧部90で行う。つまり制御信号 SAi により音響信号SAを抑圧(減衰ないし削除)し、制御信号 SBi により音響信号SBを抑圧し、制御信号 $SABi$ により両音響信号SA, SBを抑圧する。例えば信号抑圧部90内に常閉スイッチ9A, 9Bが設けられ、音源分離部80の出力端子 tA , tB が常閉スイッチ9A, 9Bを通じて、出力端子 tA' , tB' に接続され、制御信号 SAi によりスイッチ9Aが開とされ、制御信号 SBi によりスイッチ9Bが開とされ、制御信号 $SABi$ によりスイッチ9A, 9Bが共に開にされる。当然のことであるが、音源分離部80で行う分離処理するフレームの信号と、信号抑圧部90での抑圧に用いる制御信号を得るフレームの信号とは同一のものをを用いる。抑圧(制御)信号 SAi , SBi , $SABi$ の発生についてわかり易く説明する。

【0043】いま、図12に示すように音源A, Bが位置している時マイクロホンM1～M3を図に示したように配置し、ゾーンZ1～Z6を決定し、音源AとBが別個のゾーンZ3, Z4にそれぞれ位置するようにする。この時、音源AのマイクロホンM1～M3に対する距離 $SA1$, $SA2$, $SA3$ は、 $SA2 < SA3 < SA1$ となる。また、音源Bの各マイクロホンM1～M3に対する距離 $SB1$, $SB2$, $SB3$ は、 $SB3 < SB2 < SB1$ となる。

【0044】全帯域レベル検出部60の検出信号 $P(S1) \sim P(S3)$ のすべてが基準値 ThR よりも小さいとき、音源A, Bは発音、例えば発話していないと見なし、制御信号 $SABi$ により、両音響信号SA, SBを抑圧する。このとき、出力音響信号SA, SBは無音信号となる(図13の101, 102)。音源Aのみが発音しているときは、その音響信号のすべての帯域の周波数成分がマイクロホンM2へ一番大きな音圧レベル(パワー)で到達するので、このマイクロホンM2のチャンネルの合計帯域数 $\chi2$ が最も多くなる。

【0045】また、音源Bのみが発音しているときは、その音響信号のすべての帯域の周波数成分がマイクロホンM3へ一番大きな音圧レベルで到達するので、このマイクロホンM3のチャンネルの合計帯域数 $\chi3$ が最も多くなる。さらに、音源A, Bが共に発音している場合には、音響信号が最も大きな音圧レベルで到達する帯域数がマイクロホンM2とM3で拮抗する。

【0046】したがって、前記した基準値 ThP により、音響信号があるマイクロホンへ最も大きな音圧レベルで到達する合計帯域数が、当該基準値 ThP を越えた場合、当該マイクロホンが司るゾーンに音源が存在すると判定することにより、発音している音源ゾーンを検出することができる。上記の例では、音源Aのみが発音しているときは、 $\chi2$ のみが基準値 ThP を越えて、発音している音源が存在するのはマイクロホンM2が司るゾーンZ3であると検出されるので、制御信号 SBi により音声信号SBを抑制して、音響信号SAのみを出力させる(図13の103, 104)。

【0047】また、音源Bのみが発音しているときは、 $\chi3$ のみが基準値 ThP を越えて、発音している音源が存在するのは、マイクロホンM3が司るゾーンZ4であると検出されるので、制御信号 SAi により音響信号SAを抑制して、音響信号SBのみを出力させる(図13の105, 106)。さらに、音源A, Bが共に発音していて、 $\chi2$, $\chi3$ ともに基準値 ThP を越えるときは、例えば音源Aに優先度を与えて、音源Aのみが発音していると処理することができる。図13の処理手順はそうにしてある。また、 $\chi2$, $\chi3$ が共に基準値 ThP に達していない場合は、レベル $P(S1) \sim P(S3)$ が基準値 ThR を越えている限り、両音源A, Bともに発音していると判断し、制御信号 SAi , SBi , $SABi$ の何れも出力せず、音声抑圧部90では合成信号SA, SBに対する抑圧は行われない(図13の107)。

【0048】以上のようにして、音源分離部80で分離された音源信号SA, SBは、音源状態判定部70によって発音していないと判定された音源に対応するものが、信号抑圧部90で抑圧され、不要音が抑圧されるようになる。図12に示した状態に対して、図14に示すように音源CをゾーンZ6に加えた場合は、図示しない

が音源分離部80からは、音源Aに対応する信号SA、音源Bに対応する信号SBの他に、音源Cに対応する信号SCを出力する。

【0049】また、信号抑圧部90に対して、音源状態判定部70から、信号SAを抑圧する制御信号SAi、信号SBを抑圧する制御信号SBiの他に、信号SCを抑圧する制御信号SCiが出力する。また、信号SAとSBを抑圧する制御信号SABiの他に、信号SBとSCを抑圧する制御信号SBCi、信号SCとSAを抑圧する制御信号SCAi、信号SAとSBとSCの全部を抑圧する制御信号SABCiが出力する。この音源状態判定部70は、図15に示すような処理を行う。

【0050】まず、レベルP(S1)~P(S3)の全部が基準値ThRを越えていない場合は、いずれの音源A~Cも発音していないものと判断して、音源状態判定部70からSABCiを出力して、信号SA、SB、SCのいずれもが抑圧される(図15の201~202)。次に、音源A、B、Cがそれぞれ単独で発音している場合は、P(S1)~P(S3)の何れかはThRより大となり、前記した音源が2個の場合と同様に、その音源に最も近いマイクロホンのチャンネルのレベルが最も大きくなるので、そのチャンネルの帯域数 χ_1 、 χ_2 、 χ_3 のいずれかが基準値ThPを越える。そして、音源Cのみが発音している場合は、 χ_1 がThPを越え、制御信号SABiを出力して信号SA、SBが抑圧される(図15の203、204)。また、音源Aのみが発音している場合は、制御信号SBCiが出力して信号SB、SCが抑圧される。さらに、音源Aのみが発音している場合は、制御信号SBCiが出力して信号SB、SCが抑圧される(図15の205~208)。

【0051】次に、3つの音源A~Cのうちのいずれか2つが発音する場合は、発音していない音源に対応するゾーンにあるマイクロホンのレベルが最も大きくなる帯域数が、他のマイクロホンのものに比べて小さくなる。例えば、音源Cのみが発音していない場合には、マイクロホンM1のレベルが最も大きくなる帯域数 χ_1 が、他の2個のマイクロホンM2、M3の帯域数 χ_2 、 χ_3 に比べて小さくなる。

【0052】よって、予めある基準値ThQ(<ThP)を設定し、 χ_1 がその基準値ThQ以下になる場合は、マイクロホンM1とマイクロホンM3で空間を2分割したゾーンZ5、Z6の内、マイクロホンM1に近いゾーンZ6では、音源は信号を発していないと判定する。さらに、マイクロホンM1とM2で空間を2分割したゾーンZ1、Z2のうちマイクロホンM1に近いゾーンZ1では音源は信号を発していないと判定する。

【0053】すなわち、ゾーンZ1、Z6にある音源は信号を発していないと判定するのである。これらのゾーンにある音源は音源Cであることから、音源Cが信号を発していないと判定される。つまり、音源A、Bのみが

信号を発していると判定され、制御信号SCiを生成し、信号SCが抑圧される。図14に示した状態で3つの音源A~Cのうち1つのみが発音していない場合は通常は何れのマイクロホンについても最大となる帯域数 χ_1 、 χ_2 、 χ_3 は基準値ThP以下となるため、図15においてステップ203、205、207を通過し、ステップ209で、 χ_1 が基準値ThQ以下かを調べ、音源Cのみが発音していなければ、 $\chi_1 < ThQ$ となり、制御信号SCiが生成される(図15の210)。ステップ209で χ_1 がThQ以下でなければ χ_2 、 χ_3 についても同様にThQ以下であるかが順次調べられ、ThQ以下であれば音源Aのみ、又は音源Bのみが発音していないと推定され、それぞれ制御信号SAi又はSBiが抑圧される(図15の211~214)。

【0054】ステップ213で χ_3 がThQ以下でないと判定されると、音源A、B、Cは全て発音していると判定され、何れの制御信号も生成されない(図15の215)。この場合基準値ThPは $2n/3 \sim 3n/4$ 程度基準値ThQは $n/2 \sim 2n/3$ 程度、つまり例えばThPを $2n/3$ 程度にすると、ThQは $n/2$ 程度にする。

【0055】なお、以上の例では、ゾーンをZ1~Z6の6つに分けたが、図16に示すように、中心点Cpから各マイクロホン間の中点を通る点線により3つのゾーンZ1~Z3に分けても同様に音源状態を判定できる。この場合は、例えば、音源Aのみが発音している場合は、マイクロホンM2のチャンネルの帯域数 χ_2 が最も大きくなるので、そのマイクロホンM2の司るゾーンZ2に音源があると判定される。また、音源Bのみが発音している場合は χ_3 が最も大きくなり、ゾーンZ3に音源があると判定される。また、 χ_1 が予め設定した値ThQ以下である場合には、マイクロホンM1とM2およびM3とそれぞれ2分したうちのゾーンZ1にある音源は発音していないと判定する。以上の処理により、ゾーンを3分割しても、6分割のときと同様に音源の状態を判定できる。

【0056】また、基準値ThR、ThP、ThQは、全てのマイクロホンM1~M3で同一値を用いた場合で説明したが、マイクロホン毎に適宜変更してもよい。また、以上の説明では、音源が3個でマイクロホンが3個の場合についてであったが、マイクロホンの個数は音源の個数と同数以上であれば、同様に音源ゾーンを検出することができる。

【0057】例えば、音源が4個の場合には、4個のマイクロホンにより、個々のチャンネルのマイクロホンが1個の音源を司るように、図16の分割方法と同様に4個のゾーンに空間を分割する。このときの音源状態判定は、図15のステップ201~208と同様な処理により、4個全部の音源が無音か、いずれか1個が発音しているかを判定する。それらいずれでもないとき、図15

のステップ209~214と同様な処理により、4個の内の1個が無音かを判定し、1個の無音もないとき図15のステップ215と同じ処理により全部の音源が発音していると判定する。また、4個の内の3個の音源が発音しているとき(1個が無音のとき)は、そのままとしても良いが、その3個の内のより無音に近い1個を選別するには、次のようにより細かく制御する。すなわち、基準値を ThQ から ThS ($ThP > ThS > ThQ$)に換え、図15の各ステップ210, 212, 214の各々の次段に図15のステップ209~214と同様な

処理部分を設けて、3個の内から1個の無音に近い音源を判定する。

【0058】このように、音源の数が多くなるほど、図15のステップ209~214の処理内容を繰り返すことにより、無音又は無音に近い音源を2以上判定することができる。ただし、判定基準値 ThS は処理の繰り返しが増えるほど、 ThP に近付ける。以上の処理動作手順マイクロホンが4個、音源が4個の場合について図17に示すようになる。まずマイクロホン $M1 \sim M4$ より第1~第4チャンネル信号 $S1 \sim S4$ を取込み(S01)、これらチャンネル信号 $S1 \sim S4$ のレベル $P(S1) \sim P(S4)$ をそれぞれ検出し(S02)、これらレベル $P(S1) \sim P(S4)$ の何れもが基準値 ThR 以下であるかを調べ(S03)、基準値以下であれば制御信号 $SABCDi$ を生成して合成信号 $SA, SB, SC(S1)$ の出力を抑圧する(S04)。ステップS03で何れかが基準値 ThR 以下でなければ、各チャンネル信号 $S1 \sim S4$ を n 帯域に分割すると共にその各帯域のレベル $P(S1fi), P(S2fi), P(S3fi), P(S4fi)$ ($i=1, \dots, n$)を求める(S05)。各チャンネル間で同一帯域 fi のレベル中の最大のチャンネル fiM (M は1, 2, 3, 4の何れか)を各帯域について決定し(S06)、全帯域(n 個)中で $fi1, fi2, fi3, fi4$ の各合計値 $x1, x2, x3, x4$ を求める(S07)。 $x1, x2, x3, x4$ 中の最大のもの xM を求め(S08)、 xM が基準値 $ThP1$ (例えば $n/3$)以上であるかを調べ(S09)、 $ThP1$ 以上であればチャンネル M と対応して選出した音源信号、音源 A の信号であれば分離されたチャンネル M 以外の分離されたチャンネルの分離音響信号を抑圧する制御信号 $SBCDi$ を生成する(S010)。ステップS08から直ちにステップS010へ移ってもよい。

【0059】ステップS09で基準値以上でなければ xM が基準値 ThQ 以下のチャンネル M があるかを調べる(S011)。 ThQ 以下のものがなければ、全ての音源が発音しているとみなして、何れの制御信号も発生しない(S012)。ステップS011で xM が ThQ 以下のチャンネル M があれば、これと対応するチャンネル M として分離された音源信号を抑圧する制御信号 SMi を

生成する(S013)。

【0060】制御信号 SMi で抑圧された以外の分離された音源信号中の無音又は無音に近いものを抑圧するには、 S を+1し(S014) (S は予め0に初期化しておく)、 S が $M-1$ (M は音源の数)と一致したかを調べ(S015)、一致していなければ、 ThQ を $+\Delta Q$ だけ大としてステップS011に戻る(S016)。 S が $M-1$ になるまで ThQ を ThP を越えない範囲で ΔQ ずつ増加させステップS011を実行する。ステップS015で $M-1=S$ であれば、その時の ThQ 以下の各 xM の各チャンネル M と対応する分離された音源信号を抑圧する各制御信号 SMi を生成する(S013)。必要に応じてステップS015で $M-1=S$ になる前にステップS013に移ってもよい。

【0061】ステップS07で $x1 \sim x4$ を計算した後、これらで $ThP2$ (例えば $2n/3$)以上のものがあるかを調べ、あればステップS010に移り、なければステップS011に移るようにしてもよい(S017)。上述では音源分離の精度を上げるため、マイクロホン $M1 \sim M3$ のチャンネル信号 $S1 \sim S3$ の帯域間レベル差を利用して信号抑圧部90に対する制御信号を生成したが、帯域間時間差を利用して制御信号を生成することもできる。

【0062】この例を図18に、図11と対応する部分に同一符号を付けて示す。この実施例では帯域分割部41で得られた各帯域 $f1 \sim fn$ の信号 $S1(f1) \sim S1(fn)$ から到達時間差信号 $An(S1f1) \sim An(S1fn)$ が帯域別時間差検出部101で検出され、同様に帯域分割部42, 43でそれぞれ得られた各帯域の信号 $S2(f1) \sim S2(fn), S3(f1) \sim S3(fn)$ からそれぞれ到達時間差信号 $An(S2f1) \sim An(S2fn), An(S3f1) \sim An(S3fn)$ が帯域別時間差検出部102, 103で検出される。

【0063】これらの到達時間差信号を得る処理は、例えば、フーリエ変換により各帯域の信号の位相(あるいは群遅延)を算出し、同一の帯域 fi の信号 $S1(fi), S2(fi), S3(fi)$ ($i=1, 2, \dots, n$)の位相を相互に比較することで、同一音源信号の到達時間差と対応した信号を得ることができる。この場合も帯域分割部40での分割は、1つの帯域には1つの音源信号成分しか存在しないとみなせる程度に小さく行う。

【0064】この到達時間差の表現方法は、例えば、マイクロホン $M1 \sim M3$ のいずれかを基準にしてその基準マイクロホンに対する到達時間差を0に設定しておけば、他のマイクロホンに対する到達時間差はその基準マイクロホンに対して速く到達したか遅く到達したかで判定できるので、正又は負の極性を付した数値で表すことができる。この場合、基準マイクロホンを例えば $M1$ と

すると、到達時間差信号 $An(S1f1) \sim An(S1fn)$ は全て0となる。

【0065】音源状態判定部110では、コンピュータ処理により音声を発していない音源を判定する。まず、帯域別時間差検出部100により得られる到達時間差信号 $An(S1f1) \sim An(S1fn)$, $An(S2f1) \sim An(S2fn)$, $An(S3f1) \sim An(S3fn)$ を、同一の帯域の信号について相互に比較する。これにより各帯域 $f1 \sim fn$ 毎に、最も信号が速く到達するチャンネルが決定できる。

【0066】そこで、各チャンネルについて信号が最も速く到達すると判定された帯域の合計数を算出して、それをチャンネル間で比較する。この結果、この合計帯域数の値が大きいチャンネルのマイクロホンほど、音源に近いとみなすことができる。そして、あるチャンネルについて、当該合計帯域数が予め設定した基準値 ThP を越えたとき、当該のチャンネルのマイクロホンが司るゾーンに音源があると判定する。

【0067】また、この音源状態判定部110には、全帯域レベル検出部60で検出された各チャンネルのレベル $P(S1) \sim P(S3)$ も入力され、あるチャンネルのレベルが予め設定した基準値 ThR 以下の場合には、そのチャンネルのマイクロホンが司るゾーンには、音源がないと判定する。いま図12に示したように音源A、Bに対し、マイクロホンM1～M3を配置したとする。またマイクロホンM1のチャンネルに対する前記した合計帯域数を $\chi 1$ 、マイクロホンM2、M3の各チャンネルに対する合計帯域数をそれぞれ $\chi 2$, $\chi 3$ とする。

【0068】この場合も図13に示した処理手順と同様にすればよい。即ち、まず、全帯域レベル検出部60の検出信号 $P(S1) \sim P(S3)$ のすべてが基準値 ThR よりも小さいとき(101)、音源A、Bは発音していないと見なし、制御信号 $SABi$ を生成して(102)、両音源信号SA、SBを抑圧する。このとき、出力信号 SA' , SB' は無音信号となる。

【0069】音源Aのみが発音しているときは、その音源信号のすべての帯域の周波数成分がマイクロホンM2へ一番速く到達するので、このマイクロホンM2のチャンネルの合計帯域数 $\chi 2$ が最も多くなる。また、音源Bのみが発音しているときは、その音源信号のすべての帯域の周波数成分がマイクロホンM3へ一番速く到達するので、このマイクロホンM3のチャンネルの合計帯域数 $\chi 3$ が最も多くなる。

【0070】さらに、音源A、Bが共に発音している場合には、音源信号が最も速く到達する帯域数がマイクロホンM2とM3で拮抗する。したがって、前記した基準値 ThP により、音源信号があるマイクロホンへ最も速く到達する合計帯域数が、当該設定値 ThP を越えた場合、当該マイクロホンが司るゾーンに音源が存在し、その音源が発音していると判定する。

【0071】上記の例では、音源Aのみが発音しているときは、 $\chi 2$ のみが基準値 ThP を越えて(図3の103)、音響を発生している音源が存在するのはマイクロホンM2が司るゾーンZ3であると検出されるので、制御信号 SBi が生成され(104)、音響信号SBが抑制され、信号SAのみが出力される。また、音源Bのみが発音しているときは、 $\chi 3$ のみが基準値 ThP を越えて(105)、音を発生している音源が存在するのは、マイクロホンM3が司るゾーンZ4であると検出されるので、制御信号 SAi が生成され(106) 信号SAが抑制されて、信号SBのみが出力される。

【0072】この例では ThP は例えば $n/3$ 程度に設定され、音源A、Bが共に発音していて、 $\chi 2$, $\chi 3$ ともに基準値 ThP を越えることがある。この場合は図13の処理手順に示すように一方の音源、この例ではAを優先させ、音源Aへ分離信号のみを出力させることもできる。また、 $\chi 2$, $\chi 3$ が共に基準値 ThP に達していない場合は、レベル $P(S1) \sim P(S3)$ が基準値 ThR を越えている限り、両音源A、Bともに発音していると判断し、制御信号 SAi , SBi , $SABi$ は出力せず(図3の107) 音声抑圧部90では音声信号SA、SBに対する抑圧は行われない。

【0073】図12に示した状態に対して図14に示すように音源CをゾーンZ6に加えた場合、図示しないが音源分離部80からは、音源Aに対応する信号SA、音源Bに対応する信号SBの他に、音源Cに対応する信号SCが出力する。これと対応して音源状態判定部110から、信号SAを抑圧する制御信号 SAi 、信号SBを抑圧する制御信号 SBi の他に、信号SCを抑圧する制御信号 SCi が出力し、また、信号SAとSBを抑圧する制御信号 $SABi$ の他に、信号SBとSCを抑圧する制御信号 $SBCi$ 、信号SCとSAを抑圧する制御信号 $SCAi$ 、信号SA、SB、SCの全部を抑圧する制御信号 $SABCi$ が出力する。そして、この音源状態判定部110は先に述べた図15に示したと同様の処理を行う。

【0074】まず、レベル $P(S1) \sim P(S3)$ の全部が基準値 ThR を越えていない場合は、いずれの音源A～Cも発音していないものと判断して、音源状態判定部110からは $SABCi$ が出力して、信号SA、SB、SCのいずれもが抑圧される。次に、音源A、B、Cがそれぞれ単独で発音している場合には、前記した音源が2個の場合と同様に、その音源に最も近いマイクロホンのチャンネルの到達時間が最も速くなるので、そのチャンネルの帯域数 $\chi 1$, $\chi 2$, $\chi 3$ のいずれかが基準値 ThP を越える。そして、音源Cのみが発音している場合は、制御信号 $SABi$ が出力して信号SA、SBが抑圧される。また、音源Aのみが発音している場合は、制御信号 $SBCi$ が出力して信号SB、SCが抑圧される。さらに、音源Bのみが鳴っている場合は、制御信号SA

Ciが出力して信号SA, SCが抑圧される(図15の203~208)。

【0075】次に、3つの音源A~Cのうちのいずれか2つが発音している場合は、発音していない音源に対応するゾーンにあるマイクロホンの到達時間の最も速い帯域数が、他のマイクロホンのものに比べて小さくなる。例えば、音源Cのみが鳴っていない場合には、マイクロホンM1への到達時間が最も速い帯域数 $\alpha 1$ が、他の2個のマイクロホンM2, M3の帯域数 $\alpha 2$, $\alpha 3$ に比べて小さくなる。

【0076】よって、予めある基準値ThQ(<ThP)を設定し、 $\alpha 1$ がその基準値ThQ以下になる場合は、マイクロホンM1とマイクロホンM3で空間を2分割したゾーンZ5, Z6の内、マイクロホンM1に近いゾーンZ6では、音源は信号を発していないと判定し、さらに、マイクロホンM1とM2で空間を2分割したゾーンZ1, Z2のうちマイクロホンM1に近いゾーンZ1では音源は信号を発していないと判定する。

【0077】すなわち、ゾーンZ1, Z6にある音源は信号を発していないと判定するのである。これらのゾーンにある音源は音源Cであることから、音源Cが信号を発していないと判定される。つまり、音源A, Bのみが信号を発していると判定され、制御信号SCiが生成されて信号SCが抑圧される(図15の209~210)。音源Aのみ、音源Bのみがそれぞれ信号を発していないゾーンも、同様に判定される(図15の211~214)。

【0078】また、 $\alpha 1$, $\alpha 2$, $\alpha 3$ がともに基準値ThQ以下でないと判定されると、音源A, B, Cはその全てが信号を発していると判定される(図15の215)。なお、以上の例では、ゾーンをZ1~Z6の6つに分けたが、図16に示したように、3つに分けても同様に音源状態を判定できる。この場合は、例えば、音源Aのみが発音している場合は、マイクロホンM2のチャンネルの帯域数 $\alpha 2$ が最も大きくなるので、そのマイクロホンM2の可るゾーンZ2に音源があると判定される。また、音源Bのみが発音している場合は $\alpha 3$ が最も大きくなり同様にゾーンZ3に音源があると判定される。また、 $\alpha 1$ が予め設定した値ThQ以下である場合には、マイクロホンM1とM3で空間を2分したうちのゾーンZ1にある音源は発音していないと判定し、同じくマイクロホンM1とM2で空間を分割したうちのゾーンZ1にある音源は信号を発していないと判定する。以上の処理により、ゾーンを3分割しても、6分割したときと同様に音源の状態を判定できる。

【0079】以上の場合の基準値ThP, ThQの設定は、先の帯域レベルを利用する場合と同様に行えばよい。また、基準値ThR, ThP, ThQは、全てのマイクロホンM1~M3で同一値を用いた場合で説明したが、マイクロホン毎に適宜変更してもよい。また、以上

の説明では、音源が3個でマイクロホンが3個の場合についてであったが、マイクロホンの個数は音源の個数と同数以上であれば、同様に音源ゾーンを検出することができる。その処理手順は先に述べた帯域レベルを利用する場合と同様である。従って、例えば音源が4個の場合に4個の内の3個の音源が発音しているとき(1個が無音のとき)は、そのままとしても良いが、その3個の内のより無音に近い1個も選別するには、基準値をThQからThS(ThP>ThS>ThQ)に換え、図15の210, 212, 214の各々の次段に図15の209~214と同様な処理部分を設けて、3個の内から1個の無音の音源を判定することも同様である。

【0080】図17に示した処理において、そのレベルの代りに時間差を用いれば、図18に示した到達時間差を利用した不要信号の抑圧に、図17に示した処理手順も適用できる。上述においては各マイクロホンの出力チャンネル信号をまず帯域分割したが、帯域別レベルを利用する場合はまず各チャンネルのパワースペクトルを求めた後、帯域分割してもよい。その例を図19に図1, 図11と対応する部分と同一符号を付けて示し、これらと異なる部分のみを説明する。この例ではマイクロホン1, 2よりの各チャンネル信号は、パワースペクトル分解部300により、例えば高速フーリエ変換によりパワースペクトルに変換され、その後、各チャンネルごとに帯域分割部4で各帯域に分割され、各帯域ではほぼ1つの音源信号のみが主として含まれるようにして帯域別レベルを得る。この場合、音源信号選択部602へ供給する各帯域別レベルは、その原スペクトルの位相成分も供給し、音源信号合成部7で音源信号が再生できるようにする。

【0081】また各帯域別レベルは帯域別チャンネル間レベル差検出部5と音源状態判定部70とへ供給され、これらの部分で図1, 図11で説明したように処理されるその他の動作は図1又は図11の場合と同一である。図2を参照して説明した実施例において、チャンネル間時間差を用いずに、各帯域分割信号ごとに、対応帯域別チャンネル間時間差のみを用いて、何れの音源から到来したかを判定してもよい。また図5を参照して説明した実施例において、チャンネル間レベル差を用いずに、各帯域分割信号ごとに、対応帯域別チャンネル間レベル差のみを用いて、何れの音源から到来したかを判定してもよい。図5を参照した実施例におけるチャンネル間レベル差の検出は、対数レベルに変換する前のレベルを用いてもよい。図1中の帯域分割部4、図11、図18中の各帯域分割部40、図20中の帯域分割部233、図21中の帯域分割部241における各周波数帯域の分割は必ずしも同一とする必要はない。要求される精度に応じて、これらの分割数を互いに異ならせてもよい。図20中の帯域分割部233はその後の処理のために、その入力信号のパワースペクトルを先ず求め、その後、複数の周波数帯域に分割してもよい。

【0082】以下に図6～9に示したこの発明を適用した実験例を示す。図20に示す3種類の2音源信号の組み合わせにこの発明を適用し、その際に帯域分割部4で与える周波数分解能を変化させ、分離信号を物理的、及び主観的に評価した。分離処理前の混合信号は、チャンネル間時間差及びレベル差のみを計算機上で与えて加算することにより作成した。与えたチャンネル間時間差、レベル差はそれぞれ、0.47ms、2dBである。

【0083】帯域分割部4の周波数分解能は、約5Hz、10Hz、20Hz、40Hz、80Hzの5種類とした。これらの分解能で分離した信号と、原信号(OS)の計6種類の信号について評価した。なお、信号帯域は約5kHzである。定量的評価を次のように行った。混合された信号の分離が完全に行われた場合、原信号と分離信号が等しくなる。すなわち、相関係数が1となる。そこで、分離度を計る物理量として、各音について原信号と処理後の信号との相関係数を算出した。

【0084】結果を、図22に破線で示す。音声は、いずれの組み合わせについても、周波数分解能が80Hzになると相関値がかなり低くなるが、それ以外の分解能の場合は顕著な差が見られなかった。鳥の鳴き声については今回用いた周波数分解能の間に顕著な差は見られなかった。主観評価を次のように行った。

【0085】被験者は、正常な聴力を持つ20代から30代の日本人5人とした。各音源について、5種類の周波数分解能の分離音と原音をランダムにヘッドホンでダイオティックに提示し、音質について5段階で評価させた。一つの音の提示時間は約4秒間であった。結果を、図22に実線で示す。分離音S1については周波数分解能10Hzの場合が一番評価が高い。また、全ての条件に対する評価の間に有意差($\alpha < 0.05$)が存在した。分離音S2～4、6については周波数分解能20Hzの評価が最も高いが、20Hzと10Hzとの間には有意差はなかった。また、20Hzの音と5Hz、40Hz、80Hzの間にはそれぞれ有意差が存在した。こ*

	男(S1)	女(S2)
検出率	99%	93%

図6～9に示した基本方法と図11に示した改良方法とのそれぞれで分離した音をランダムにヘッドホンでダイオティックに提示し、雑音の交じり具合の少なさと不連続感の少なさについて評価させた。用いた分離音は前記S1～S4であり、被験者は正常な聴力を持つ20代から30代の日本人5名である。一つの音の提示時間は約4秒間、各音の試行回数は3回である。その結果、雑音の交じり具合が少ないと評価した率は改良方法が91.7%、基本方法は8.3%で、改良方法が少ないと判断した回答が格段と多かった。一方不連続感が少ないについては改良方法は20.0%、基本方法が80.0%で基本方法の方が少ないと判断する回答が多かったが、改良方法との間に有意な差は見られなかった。

*これらの結果から、音声については分離する組み合わせの種類によらず、最適な周波数分解能が存在することが分かった。この実験の場合は20Hzもしくは10Hz程度が最適な値である。分離音S5(鳥の鳴き声)については40Hzの場合が最も評価が高いが有意差は40Hzと5Hz、20Hzと5Hzの間にしか存在しなかった。なお、いずれの場合についても、分離処理後の音と原音の間には有意差が存在した。

【0086】図21、図23にこの発明の効果を示す。

図21は、分離処理前の男声と女声の混合音声のスペクトル201とこの発明による分離処理後の男声S1、女声S2の各スペクトル202、203を表す。図23は、分離処理前の男声S1、女声S2の各原音声の各波形をA、Bに、混合音声波形をCに、分離処理後の男声S1、女声S2の各波形をD、Eにそれぞれ示す。図21からは、不要な成分が抑圧されていることが分かる。さらに、図23からは、分離処理後の音声は原音声と同程度の品質で復元されていることが分かる。

【0087】帯域分割の分解能は音声の場合、10～20Hz程度が好ましく、5Hz以下、50Hz以上は好ましくない。帯域分割の手法はフーリエ変換に限らず、帯域フィルタにより分割してもよい。次に図11に示したレベル差を利用して音源状態を判定して信号抑圧部90で信号抑圧を行う場合の実験例を示す。2個のマイクロホンを用い、2つの音源A、Bをダミーヘッドから距離1.5m、角度差90度(2つのマイクロホンの中点に対し右45度、左45度)の位置に置き、同一の音圧レベルで、残響時間0.2s(500Hz)の可変残響室内で収音した。用いた混合音と分離音の組み合わせは図22中のS1～S4である。

【0088】分離音声S1～S4について、無音と判定されたフレームの個数と、原音の無音フレームの個数の比率を算した。その結果は次の通り90%以上正しく検出された。

女声1(S3)	女声2(S4)
92%	95%

【0089】次に分離性能を相対評価を行うため、以下の5種類の音の分離度の比較を主観評価により行った。

(1) 原音

(2) 基本法(計算機)：チャンネル間時間差(0.47ms)、レベル差(2dB)を与えて計算機上で加算した混合信号を、基本方法で分離した音。

(3) 改良法(実環境)：先の無音区間検出率の実験に用いた条件で収音した混合音を改良方法で分離した音。

(4) 基本法(実環境)：先の無音区間検出率の実験に用いた条件で収音した混合音を基本方法で分離した音。

(5) 混合音：先の無音区間検出率の実験に用いた条件で収音した混合音。

【0090】図20中の最初の2つの混合音に対し、

“原音”上記(1)～(4)の方法で処理した音、“混合音”の計20種類をランダムにヘッドホンでダイオティックに提示し、分離度について7段階で評価させた。つまり「最も分離されている」を7点、「最も分離されていない」を1点とした。被験者、音の提示時間及び試行回数は、前記雑音の交じり具合の少なさの評価の場合と同一である。

【0091】この結果を図24中で、全音源(S0)をAに、男声(S1)をBに、女声(S2)をCに、女声1(S3)をDに、女声2(S4)をEにそれぞれ示す。全音源について分析した結果(S0)と、音源の種類毎に分析した結果(S1)～(S4)とは、ほぼ同じ傾向を示した。S0～S4全ての場合について、“(1)原音”、“(2)基本法(計算機)”、“(3)改良法(実環境)”、“(4)基本法(実環境)”、“(5)混合音”の順に分離精度が高い。つまり実環境では改良方法の方が基本方法より優れている。

【0092】

【発明の効果】以上述べたようにこの発明によれば複数のマイクロホンからの各チャンネル信号を、主な成分が1つの音源信号の成分のみからなる程度に複数の帯域に分割し、これら各同一帯域について、レベル、到達時間を検出し、これらから、各帯域ごとに何れの音源信号かを判定分離することにより、各音源信号を正しく分離することができ、しかも実時間での処理が可能である。

【0093】特に発音していない音源を検出し、その成分を抑圧することにより、部屋内のような回り込みや、残響がある場所でも、正確に分離することができる。

【図面の簡単な説明】

【図1】この発明の音源分離装置の実施例の機能構成を示すブロック図。

【図2】この発明の音源分離方法の実施例の処理手順を示す流れ図。

【図3】図2中のチャンネル間時間差 $\Delta\tau_1$ 、 $\Delta\tau_2$ を求める処理手順の例を示す流れ図。

【図4】A、Bはそれぞれ二つの音源信号のスペクトルの例を示す図である。

【図5】この発明の音源分離方法で、チャンネル間レベル差を利用して音源分離を行う実施例の処理手順を示す流れ図。

【図6】この発明音源分離方法で、チャンネル間レベル差

と、チャンネル間到達時間差を利用する実施例の処理手順の一部を示す流れ図。

【図7】図6中のステップS08の続きを示す流れ図。

【図8】図6中のステップS09の続きを示す流れ図。

【図9】図6中のステップS10、図7、図8中のステップS20、S30の続きを示す流れ図。

【図10】周波数帯域が異なる音源信号を分離する実施例の機能構成を示すブロック図。

【図11】レベル差を利用して不要音源信号を抑圧する構成を付加したこの発明の音源分離装置の実施例の機能構成を示すブロック図。

【図12】3つのマイクロホンとその受けもつゾーンと、2つの音源の配置例を示す図。

【図13】発音している音源が1つの場合の音源ゾーンの検出と、抑圧制御信号の生成処理手順の例を示す流れ図。

【図14】3つのマイクロホンと、その受けもつゾーンと、3つの音源の配置例を示す図。

【図15】音源が3つの場合の発音音源のゾーン検出と、抑圧制御信号の生成処理手順の例を示す流れ図。

【図16】3つのマイクロホンによりゾーンを3つに分割した例と、音源の配置例を示す図。

【図17】この発明の音源分離装置において、発音していない合成音源信号を抑圧する制御信号を生成するための処理手順の例を示す流れ図。

【図18】到達時間差を利用して不要音源信号を抑圧する構成を付加したこの発明の音源分離装置の実施例の機能構成を示すブロック図。

【図19】この発明音源分離装置で、パワースペクトルを求めた後、帯域分割を行う場合の実施例の機能構成を示すブロック図。

【図20】この発明の実験に用いた音源の種類を示す図。

【図21】図6～図9に示した実施例の方法による処理前と、処理後の音声スペクトルを示す図。

【図22】図6～図9に示した実施例の方法を用いた主観評価実験の結果を示す図。

【図23】図6～図9に示した実施例の方法により処理した処理後の音声波形と、その原音声波形を示す図。

【図24】図6～図9に示した音源分離方法と図11に示した音源分離装置とについての実験結果を示す図。

【図1】

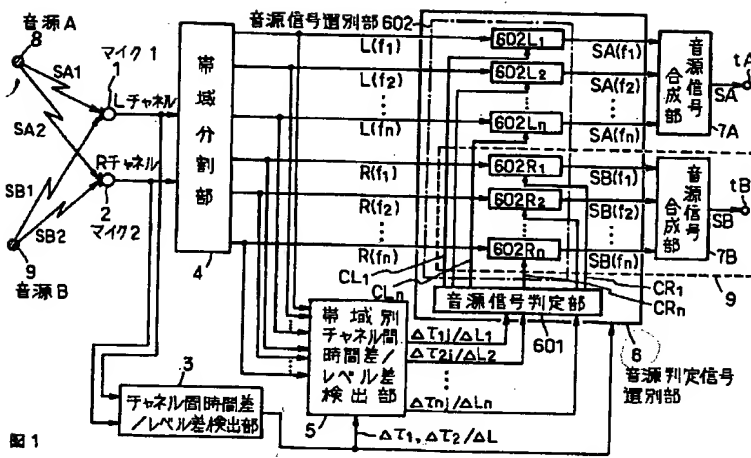


図1

【図4】

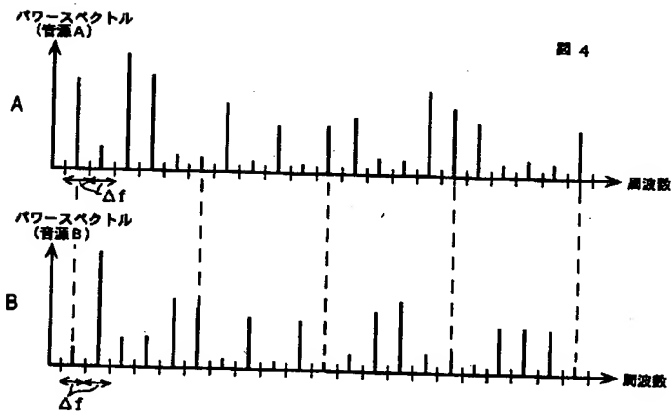


図4

【図13】

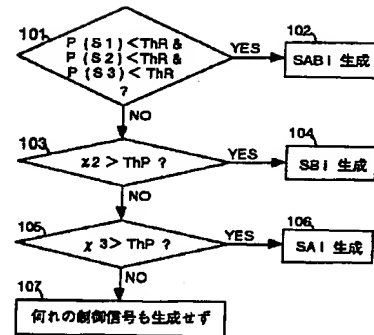


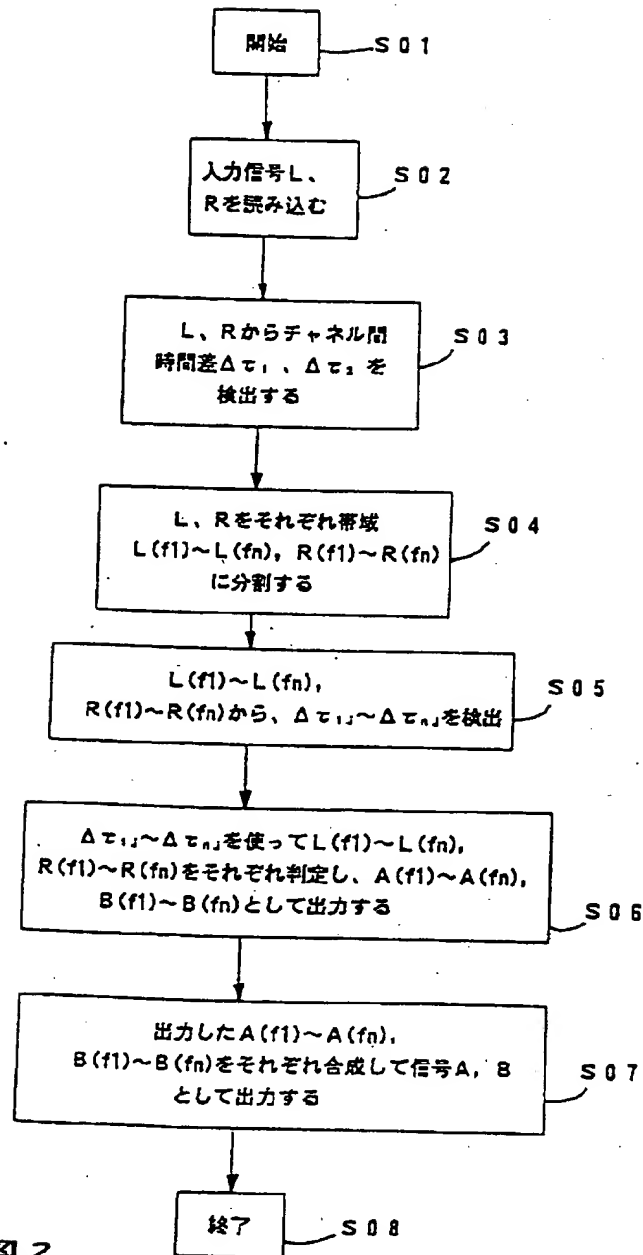
図13

【図20】

混合信号	分離信号
男声+女声	男声(S1)
	女声(S2)
女声1+女声2 (同一発話者 (発話内容は異なる))	女声1(S3)
	女声2(S4)
おんどのりのおき声+女声	おんどのりのおき声(S5)
	女声(S6)

図 20

【図2】



【図3】

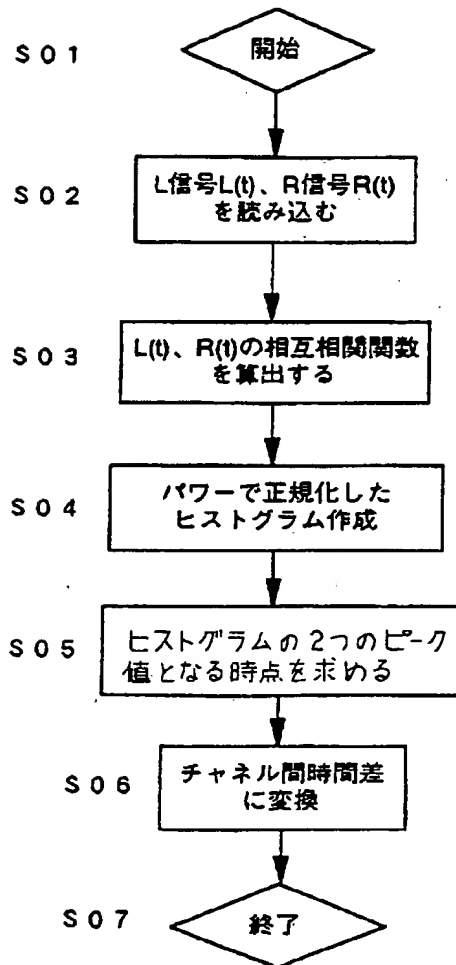


図3

【図14】

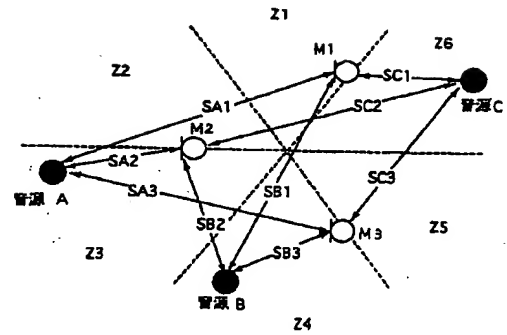


図14

【図15】

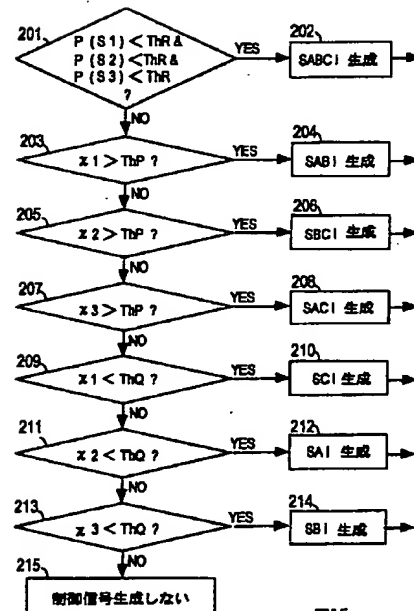


図15

【図5】

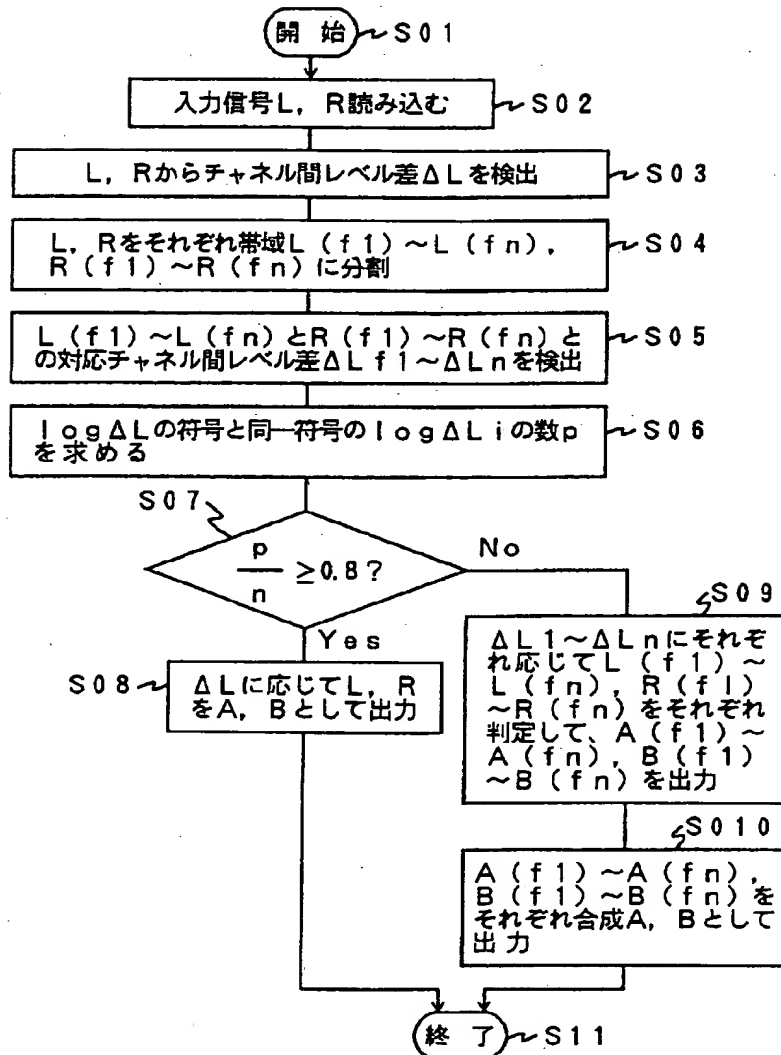


図 5

【図6】

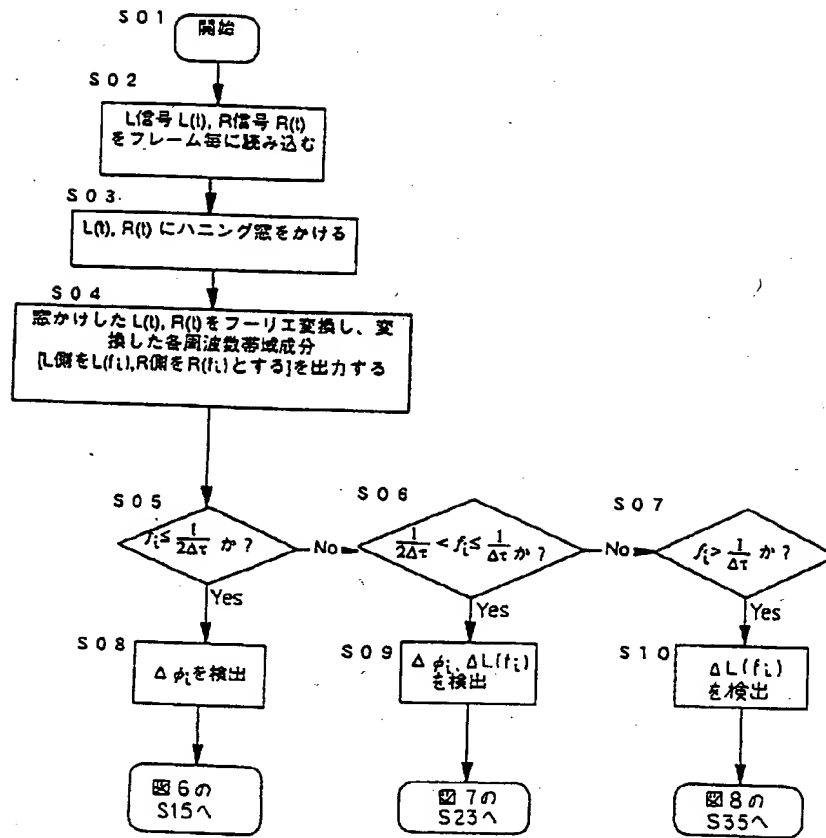
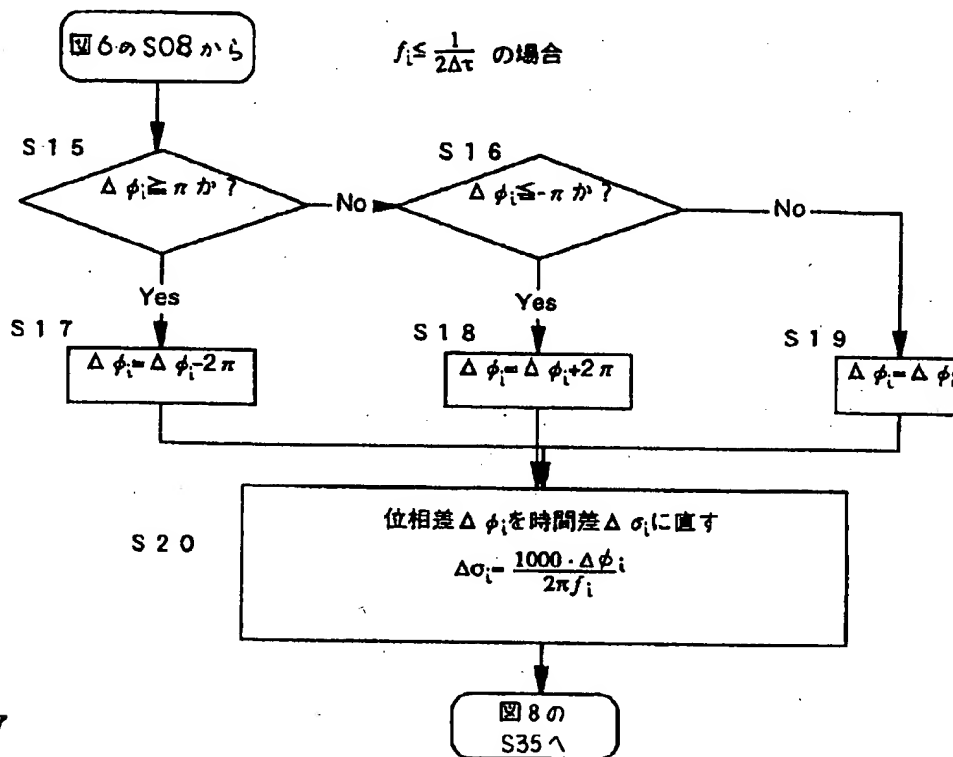
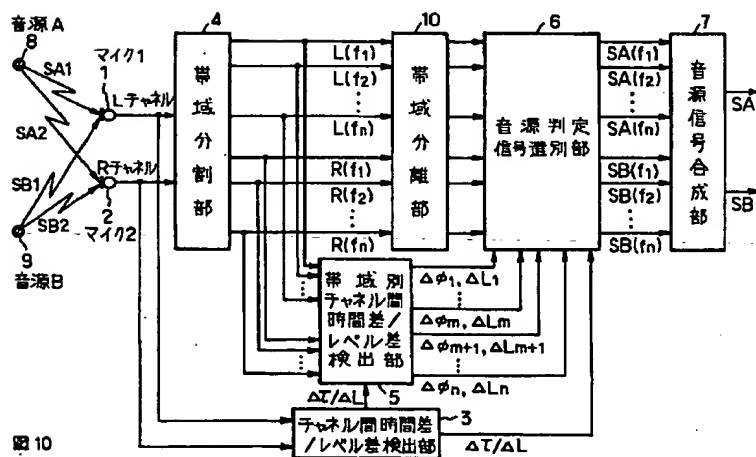


図6

【図7】



【図10】



【図8】

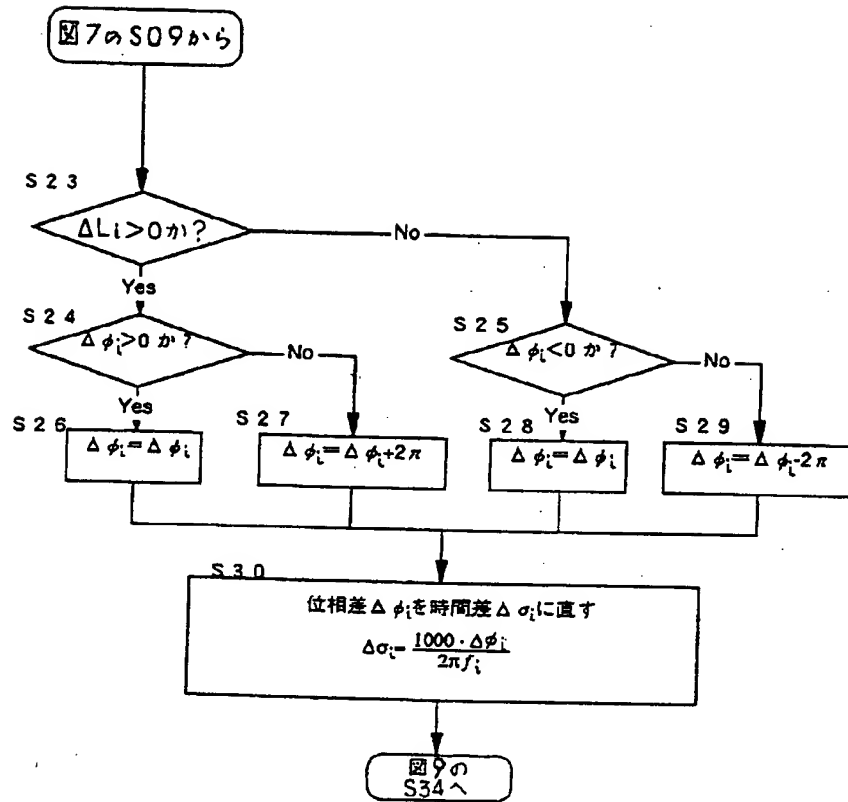
 $\frac{1}{2\Delta\tau} < f_i \leq \frac{1}{\Delta\tau}$ の場合


図 8

【図9】

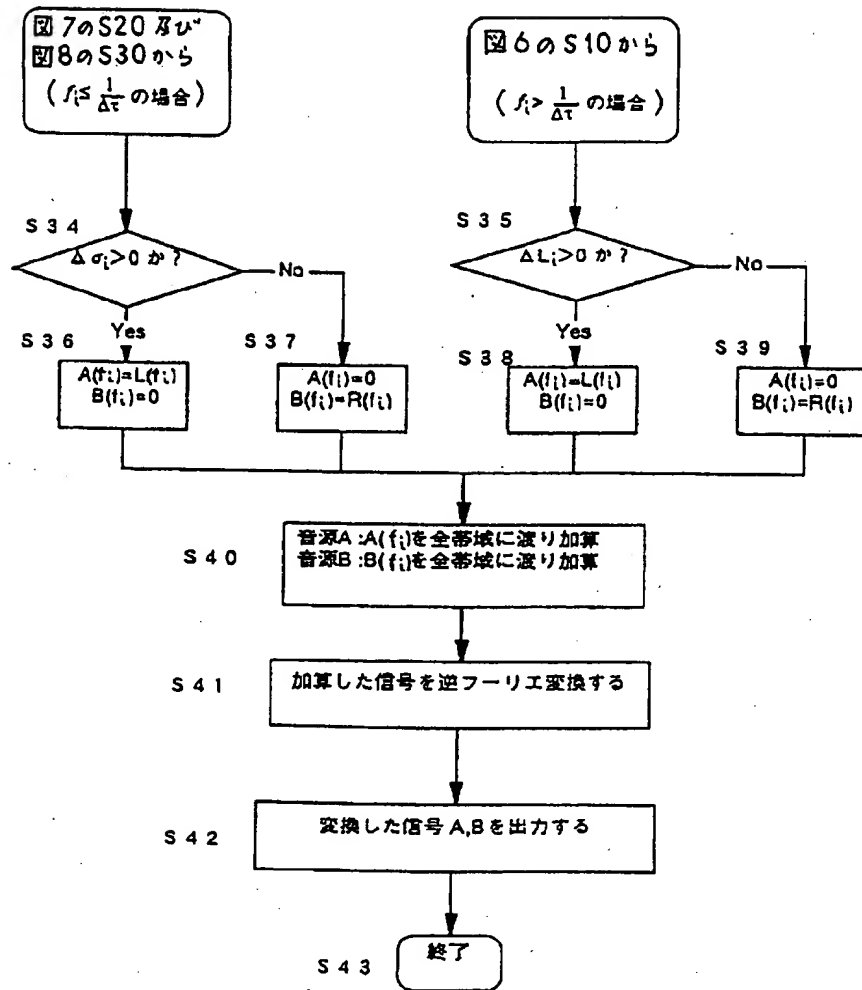


図 9

【図11】

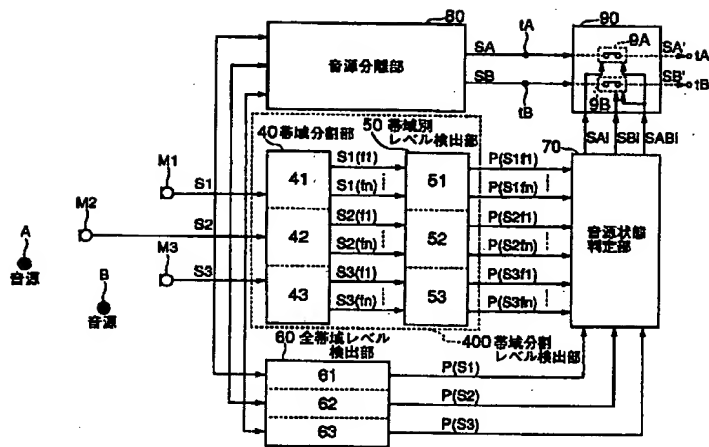


図 11

【図12】

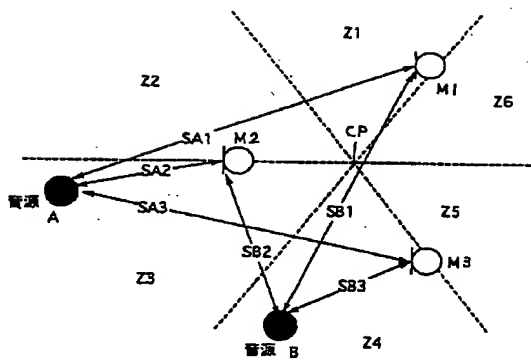


図 12

【図16】

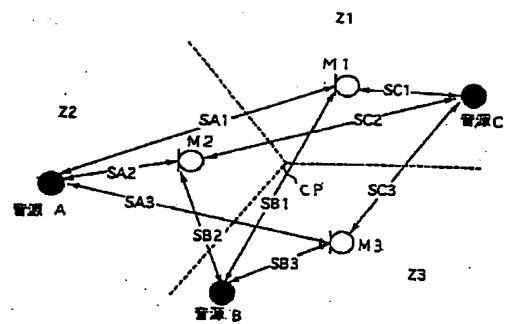


図 16

【図17】

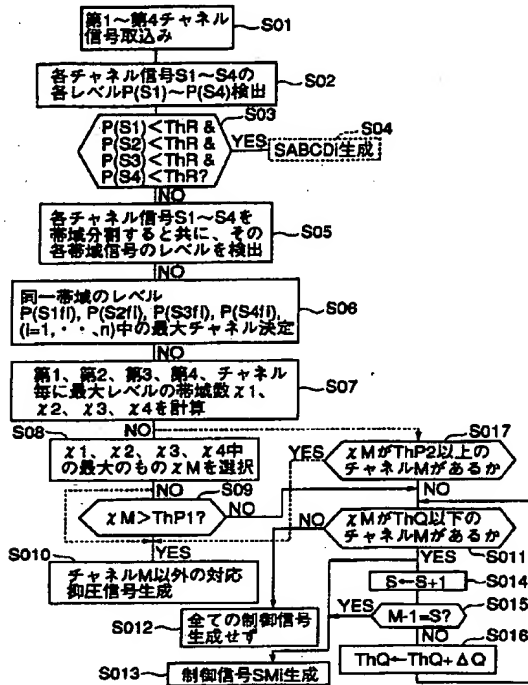


図 17

【図21】

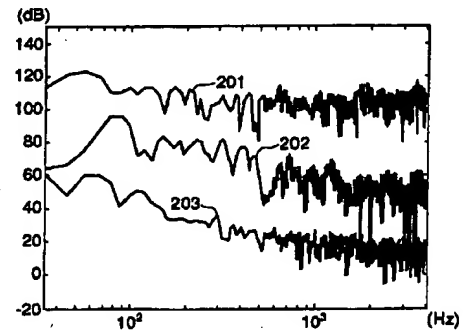


図 21

【図18】

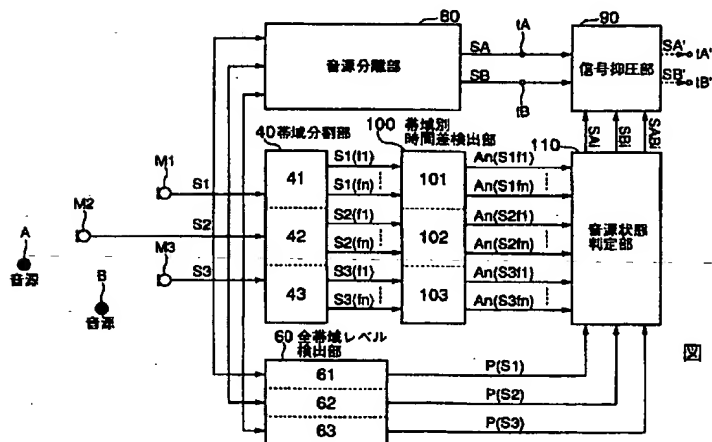
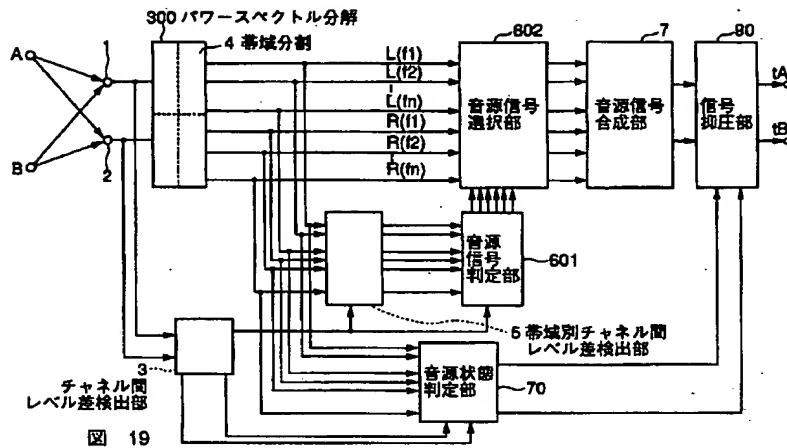


図 18

【図19】



【図22】

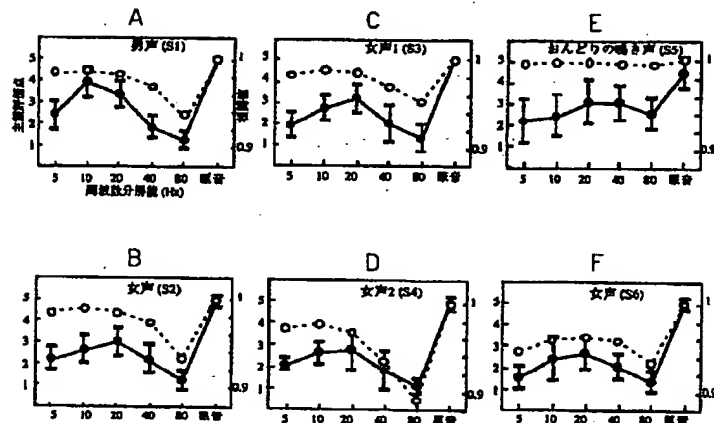


図 22

【図23】

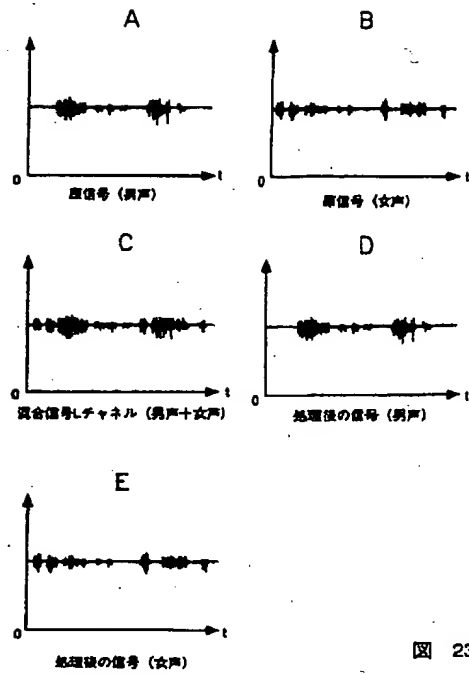


図 23

【図24】

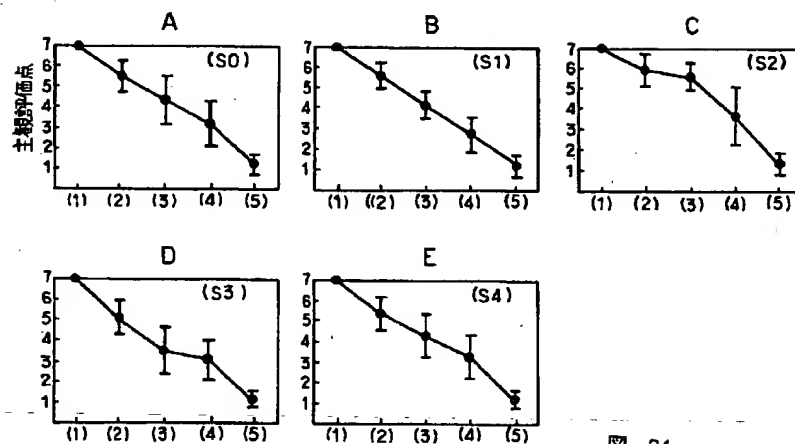


図 24

【手続補正書】

【提出日】平成9年9月25日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】請求項20

【補正方法】変更

【補正内容】

【請求項20】 請求項19の方法において、

上記第1判定過程で、第1基準値を越えるものがないと判定されると、上記最も速く到達する帯域の数が、上記第1基準値よりも小さい第2基準値よりも小さいか否かを判定する第2判定過程と、
 その第2判定過程で、第2基準値よりも小さいと判定されると、その小さいと判定された最も速い到達時間の帯域数と対応するチャンネルのマイクロホン位置から、発音していない1個の音源として検出する過程とを有することを特徴とする音源分離方法。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】請求項21

【補正方法】変更

【補正内容】

【請求項21】 請求項16又は20の方法において、音源が4個以上の場合で、上記第2判定過程で、第2基準値よりも小さいと判定されると、上記第2基準値を上記第1基準値を越えない範囲内で、順次大きくして、上記第2判定過程と同じ判定を、(M-2)回以内、Mは音源の数、繰返す過程を有することを特徴とする音源分離方法。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】請求項33

【補正方法】変更

【補正内容】

【請求項33】 請求項32の装置において、
 上記第1判定手段で、第1基準値を越えるものがないと判定されると、上記最もレベルが大きい帯域の数が、上*

*記第1基準値よりも小さい第2基準値以下か否かを判定する第2判定手段と、

その第2判定手段で、第2基準値よりも小さいと判定されると、その小さいと判定された最もレベルが大きい帯域の数と対応するチャンネルのマイクロホン位置から、発音していない1個の音源として検出する手段とを有することを特徴とする音源分離装置。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】請求項35

【補正方法】変更

【補正内容】

【請求項35】 請求項34の装置において、

上記音源状態判定手段は、上記各帯域別到達時間差比較で最も速く音源信号が到達したチャンネルを帯域ごとに決定する手段と、

各チャンネルごとに最も速く到達した帯域の数が第1基準値を越えるか否かを判定する第1判定手段と、

その第1判定手段が第1基準値を越えると判定すると、その越えた最も速く到達した帯域数と対応するチャンネルのマイクロホン位置から発音している1個の音源を推定する手段と、

その推定された音源以外の音源を発音していないものとして検出する手段とを有することを特徴とする音源分離装置。

【手続補正5】

【補正対象書類名】明細書

【補正対象項目名】請求項40

【補正方法】変更

【補正内容】

【請求項40】 請求項39の記録媒体において、

上記チャンネル時間差検出過程は各出力チャンネル信号間の相互相関を求め、相互相関の各ピークとなる、その出力チャンネル信号間の各時間差を上記各チャンネル間時間差として求めることを特徴とする記録媒体。

フロントページの続き

(51)Int.Cl.⁶

H04S 7/00

識別記号

FI

H04S 7/00

Z

(72)発明者 西野 豊

東京都新宿区西新宿三丁目19番2号 日本
 電信電話株式会社内